

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:14:52 ; Search time 73 Seconds
(without alignments)
848.788 Million cell updates/sec

Title: US-09-388-090-4

Perfect score: 2353

Sequence: 1 VPKKYYQVFAALACALLAG.....ERAGLRHGDDEILAVRASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

A.GeneSeq.101002.*
1: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneSeq/geneSeq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2353	100.0	465	21	NGSP polypeptide o
2	2317	98.5	499	21	Neisseria gonorrhoe
3	2254	95.8	499	21	Neisseria meningit
4	2248	95.5	499	21	Neisseria meningit
5	2244	95.4	499	21	Neisseria meningit
6	2234.5	95.0	499	21	Neisseria meningit
7	2202	93.6	499	21	Neisseria meningit
8	2197	93.4	499	21	Neisseria meningit
9	2130.5	90.5	475	21	Neisseria meningit
10	2092	88.9	414	21	NGSP polypeptide o

11	2002	85.1	448	21	AAV70409
12	1832	77.9	370	21	AAV52996
13	962.5	40.9	460	20	AAV29294
14	731.5	31.1	503	17	AAV77434
15	705	30.0	474	22	AAV78605
16	685	29.1	491	20	AAV37309
17	668.5	28.4	488	13	ABR90581
18	633.5	26.9	475	12	AAV14625
19	632	26.9	463	17	AAV26630
20	629	26.7	437	21	AAV18340
21	627.5	26.7	455	22	AAV98956
22	624	26.5	437	21	AAV11837
23	624	26.5	437	21	AAV11838
24	623	26.4	437	21	AAV11836
25	622	26.4	437	21	AAV11835
26	603.5	25.6	491	19	AAV56771
27	573.5	24.4	443	20	AAV98950
28	565.5	24.0	451	20	AAV89977
29	559.5	23.8	387	19	AAV8445
30	542.5	23.1	433	21	AAV59353
31	542.5	23.1	433	21	AAV59354
32	533	22.7	433	21	AAV59354
33	453	19.3	397	21	AAV81772
34	449.5	19.1	320	21	AAV80962
35	448	19.0	320	21	AAV80958
36	448	19.0	382	21	AAV80958
37	448	19.0	394	21	AAV51193
38	448	19.0	451	21	AAV80957
39	448	19.0	480	18	AAV2849
40	448	19.0	480	21	AAV80956
41	448	19.0	480	21	AAV5191
42	448	19.0	517	23	ABP41178
43	447.5	19.0	355	22	AAV98947
44	445	18.9	480	21	AAV80964
45	443.5	18.8	405	21	AAV58844

ALIGNMENTS

RESULT 1
ID AAV83150 standard; Protein; 465 AA.
AAV83150;
24-JUL-2000 (first entry)
NGSP polypeptide of Neisseria gonorrhoeae.
DE NGSP polypeptide of Neisseria gonorrhoeae.
XX NGSP: polypeptide; peptide; vaccine; immune response; antibody;
KW cellular matrix; adherence domain; ligand; detection; diagnosis;
screening; probe; primer; prophylaxis; therapy.
XX Neisseria gonorrhoeae.
OS Neisseria gonorrhoeae.
XX WO200012133-A1.
PN 09-MAR-2000.
PD 01-SEP-1999; 99WO-US20070.
PF 01-SEP-1999; 98US-0098685.
PR 01-SEP-1998; 98US-0098685.
PA (ANTE-) ANTEX BIOLOGICS INC.
XX Jackson WJ, Harris AM;
PI WPI; 2000-237782/20.
DR N-PSDB; AA293414.
XX Non-cytosolic NGSP polypeptide and polynucleotide sequence from
PT Neisseria useful for diagnosis, prevention or treatment of Neisseria

Neisseria meningit
Neisseria meningit
Protein encoded by
Heat shock protein
Lawsonia intracell
Protein involved i
Chlamydia pneumon
S. typhimurium Htr
H. influenzae SB33
Haemophilus influe
E. coli growth and
Haemophilus influe
Haemophilus influe
Haemophilus influe
Haemophilus influe
Becherichia coli s
Antigen 1 from clu
Protein encoded by
H. pylori GHPO 536
BASB01 protein se
BASB01 protein se
Mycobacterium tube
Streptococcus pneu
Human HtrA (high t
Human HtrA (high t
C-terminal domain
Human HtrA (high t
Osteoblast like ce
Human HtrA (high t
Human serine prote
Human ovarian anti
E. coli growth and
Human HtrA (high t
Breast and ovarian

PT infections
 XX Claim 5; Page 61-62; 68pp; English.
 PS
 XX The NGSP polypeptide of *N. gonorrhoeae* has conserved Arg-Gly-Asp and
 CC Arg-Gly-Asn motifs near the C-terminus which function as adherence
 CC domains for extracellular matrix proteins. Using the NGSP polypeptide
 CC as a vaccine produces antibodies which inhibit binding of *N.*
 CC gonorrhoeae to the host's cellular matrix reducing attachment and/or
 CC subsequent invasion. The NGSP polypeptide and its peptide fragments
 CC can be used to immunise an animal and produce an immune response.
 CC They can also be used as ligands to detect antibodies elicited in
 CC response to *Neisseria* infections and also as antigens or immunogens
 CC for inducing *Neisseria*-specific antibodies which are useful in
 CC immunoassays to detect *Neisseria* in biological specimens. Nucleotides
 CC encoding NGSP or its fragments can be used as probes to identify
 CC *Neisseria* in biological specimens by hybridization or polymerase
 CC chain reaction amplification. The NGSP polypeptide can also be used
 CC in screening assays to identify agents and compounds which useful as
 CC diagnostic, prophylactic or therapeutic agents against *Neisseria*
 CC infection.

Sequence 465 AA;
 Query Match 100.0%; Score 2353; DB 21; Length 465;
 Best Local Similarity 100.0%; Pred. No. 1,6e-190;
 Matches 465; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 VFKKYQYFALAAACALLAGECKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 1 VFKKYQYFALAAACALLAGECKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 121 GGLNFGSGFIISKNGYILLTNTHVAVGMSIKVLNDKREYTKLIGSDVQSDVALLKIDA 180
 121 GGLNFGSGFIISKNGYILLTNTHVAVGMSIKVLNDKREYTKLIGSDVQSDVALLKIDA 180
 121 GGLNFGSGFIISKNGYILLTNTHVAVGMSIKVLNDKREYTKLIGSDVQSDVALLKIDA 180
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 241 INPGNSGGLPLNLKQVVGINSQIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300
 241 INPGNSGGLPLNLKQVVGINSQIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300
 301 LGVLIQEVSYGLAQSPGLDASGALIAKILPGSPAERAGLQAGDIVSLDGEIRSSGDL 360
 301 LGVLIQEVSYGLAQSPGLDASGALIAKILPGSPAERAGLQAGDIVSLDGEIRSSGDL 360
 361 PVMVGAITPEKEVSLGVRKGEETITAKIGNAAEHTGASGKTDBAPTEOOSGTFVS 420
 361 PVMVGAITPEKEVSLGVRKGEETITAKIGNAAEHTGASGKTDBAPTEOOSGTFVS 420
 421 AGTTLQTHDSSGKLTVVRVSDAERAGLRHGEILLAVRASPRQ 465
 421 AGTTLQTHDSSGKLTVVRVSDAERAGLRHGEILLAVRASPRQ 465

RESULT 2
 ID AAY75748 standard; Protein: 499 AA.
 XX AAY75748;
 XX 21-MAR-2000 (first entry)
 DE *Neisseria gonorrhoeae* ORF 986 protein sequence SEQ ID NO:2968.
 XX *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
 KW antigenic; diagnostic; immunogenic; infection; meningitis; septicemia;

KW antibacterial; gene therapy.
 XX
 OS *Neisseria gonorrhoeae*.
 XX
 XX WO9957280-A2.
 XX
 PD 11-NOV-1999.
 XX
 PF 30-APR-1999; 99WO-US09346.
 XX
 PR 01-MAY-1998; 98US-0083758.
 PR 31-JUL-1998; 98US-0094869.
 PR 02-SEP-1998; 98US-0098994.
 PR 02-SEP-1998; 98US-0099062.
 PR 09-OCT-1998; 98US-0103749.
 PR 09-OCT-1998; 98US-0103794.
 PR 09-OCT-1998; 98US-0103796.
 PR 25-FEB-1999; 99US-0121528.
 XX
 PA (CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.
 XX
 PI Frazer C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
 PI Petersen J, Pizza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
 PI Tettelin H, Venter JC;
 XX
 DR WPI: 2000-062150/05.
 DR N-P8DB; AA254510.
 XX
 PS Claim 2; Page 1390; 1453pp; English.
 PT Novel *Neisseria* polypeptides predicted to be useful antigens for
 PT vaccines and diagnostics -
 XX
 CC AA253015 to AA254536, AA254577 to AA254615, and AAY74253 to AAY75941
 CC represent novel *Neisseria meningitidis* and *N. gonorrhoeae* polynucleotides
 CC and polypeptides. AA254537 to AA254576 and AA254616 to AA25473 represent
 CC PCR primers used in the exemplification of the present invention. The
 CC polypeptides, the polynucleotides, antibodies and compositions of
 CC the invention can be used as vaccine, as diagnostic reagents, and as
 CC immunogenic compositions. The polypeptides can be used in the
 CC manufacture of medicaments for treating or preventing infection due to
 CC *Neisseria* bacteria (e.g. meningitis and septicemia), to detect the
 CC presence of *Neisseria* bacteria, or to raise antibodies. They may also
 CC be used to screen for agonists or antagonists, which may themselves
 CC have use as antibacterial agents. The polynucleotides of the invention
 CC may also be used in gene therapy protocols.
 XX

Sequence 499 AA;
 Query Match 98.5%; Score 2317; DB 21; Length 499;
 Best Local Similarity 99.1%; Pred. No. 2e-187;
 Matches 459; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

1 VFKKYQYFALAAACALLAGECKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 1 VFKKYQYFALAAACALLAGECKAGSFFGADKKEASFYERIEHTKDDGSVSM LLPFAQL 60
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 61 VOSEGPAAVNIQAAPAPRTONGSNAETDSDPLADSDPFYEFKRLVPMMEIPEEADD 120
 121 GGLNFGSGFIISKNGYILLTNTHVAVGMSIKVLNDKREYTKLIGSDVQSDVALLKIDA 180
 121 GGLNFGSGFIISKNGYILLTNTHVAVGMSIKVLNDKREYTKLIGSDVQSDVALLKIDA 180
 121 GGLNFGSGFIISKNGYILLTNTHVAVGMSIKVLNDKREYTKLIGSDVQSDVALLKIDA 180
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 181 TEELPVVXIGNPKNLKPEEWAAIAGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTDVA 240
 241 INPGNSGGLPLNLKQVVGINSQIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300
 241 INPGNSGGLPLNLKQVVGINSQIYSRSGGFMGISFPAIPDIVAMNVAEQLKNTGKVRQ 300

QY 301 LGVITIEVSYGLAOSFGDLKASGALIAKILPGSPAPERAGLQAGDIVLSLDGGEIRSSGDL 360
 Db 301 LGVITIEVSYGLAOSFGDLKASGALIAKILPGSPAPERAGLQAGDIVLSLDGGEIRSSGDL 360
 QY 361 PVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGASSTKDEAPYTEQSGTSFVES 420
 Db 361 PVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGASSTKDEAPYTEQSGTSFVES 420
 QY 421 AGITLQHTDSSGKHLVVRVSDAERAGLRHGDDEILAVRASP 463
 Db 421 AGITLQHTDSSGKHLVVRVSDAERAGLRHGDDEILAVGQVP 463

RESULT 3
 ID AAY52995 standard; Protein: 499 AA.
 AC AAY52995;

21-FEB-2000 (first entry)

Neisseria meningitidis strain H44/76 BASB013 protein sequence.
 DE Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
 XX Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
 KW antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
 KM invasive bacterial disease; antibacterial.
 XX

Neisseria meningitidis.

WO955872-A1.

04-NOV-1999.

20-APR-1999; 99WO-EP02765.

23-APR-1998; 98GB-0008734.

(SMIK) SMITHKLINE BEECHAM BIOLOGICALS.

Ruelle J;

WPI; 2000-052809/04.

N-PSDB; AAZ33307.

Novel polynucleotides and polypeptides from Neisseria meningitis used to prepare vaccines against bacterial infections

Claim 3; Page 79-81; 94pp; English.

The present sequence represents a BASB013 polypeptide isolated from Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be employed as research reagents and material for the discovery of treatments and diagnostics for diseases, particularly human diseases. They can be used for diagnosis of disease, staging of disease, or determining response of an infectious organism to drugs. The polynucleotides may be used as a source for hybridisation probes, and for screening of genetic mutations, serotype, organism or strain identification, identification of mutation in BASB013 sequences, and as components of arrays which are useful for diagnostic and prognostic purposes. The polypeptides can be used to produce antibodies. The polypeptides can also be used in vaccine formulations, and to identify antigens and antagonists. The polypeptides, antibodies, agonists and antagonists (which are bacteriostatic) are used for the treatment and prevention of diseases such as upper respiratory tract infection, invasive bacterial diseases such as bacteraemia and meningitis, and for the development and screening of antibacterial drugs. They are also used in the prevention of adhesion of bacteria to eukaryotic matrix proteins on in-dwelling devices, or to extracellular proteins on wounds, and to thus prevent tissue damage and/or block the normal progression of pathogens in infections initiated other than by the implantation of in-dwelling devices or by other surgical techniques.

SEQ Sequence 499 AA;

Query Match 95.8%; Score 2254; DB 21; Length 499;
 Best Local Similarity 96.1%; Pred. No. 4.4e-182;
 Matches 445; Conservative 6; Mismatches 12; Indels 0; Gaps 0;

QY 1 VFKKYQYFALAAALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGSVSMILPDPFQOL 60
 Db 1 MFKKYQYFALAAALCAALLAGCEKAGSFFGADKKEASFVERIEHTKDDGSVSMILPDPFQOL 60
 QY 61 VQSEGPVAVNIQAAPAPRTQNGSGNAENDSDPIANDPFEFFKXLVNMPETIPOEABDD 120
 Db 61 VQSEGPVAVNIQAAPAPRTQNGSGNAENDSDPIANDPFEFFKXLVNMPETIPOEABDD 120
 QY 121 GGLNFGSGFIISKNGIYILNTHVAVGMSIKVLLNDKREYTKLIGSDVQSDVALIKTDA 180
 Db 121 GGLNFGSGFIISKNGIYILNTHVAVGMSIKVLLNDKREYTKLIGSDVQSDVALIKTDA 180
 QY 181 TEELPVVKIGNPKNLKPGEWVAATGAPGFDNSVTAGIYSAKGRSLPNESYTPFIQTDVA 240
 Db 181 TEELPVVKIGNPKNLKPGEWVAATGAPGFDNSVTAGIYSAKGRSLPNESYTPFIQTDVA 240
 QY 241 INPNSGGFLFKLKGQVNSINSQIYSRSGFMGISFAIPIDVAMNVAEQLKNTGKVQRQ 300
 Db 241 INPNSGGFLFKLKGQVNSINSQIYSRSGFMGISFAIPIDVAMNVAEQLKNTGKVQRQ 300
 QY 301 LGVITIEVSYGLAOSFGDLKASGALIAKILPGSPAPERAGLQAGDIVLSLDGGEIRSSGDL 360
 Db 301 LGVITIEVSYGLAOSFGDLKASGALIAKILPGSPAPERAGLQAGDIVLSLDGGEIRSSGDL 360
 QY 361 PVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGASSTKDEAPYTEQSGTSFVES 420
 Db 361 PVMVGAITPGKEVSLGVWRKGEIITIKAKLGNAAEHTGASSTKDEAPYTEQSGTSFVES 420
 QY 421 AGITLQHTDSSGKHLVVRVSDAERAGLRHGDDEILAVRASP 463
 Db 421 AGITLQHTDSSGKHLVVRVSDAERAGLRHGDDEILAVGQVP 463

RESULT 4

AAY75749 standard; Protein: 499 AA.

AAY75749;

21-MAR-2000 (first entry)

Neisseria meningitidis ORF 986 protein sequence SEQ ID NO:2970.

Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
 KW antigenic; diagnosis; immunogenic; infection; meningitis; septicemia;
 KM antibacterial; gene therapy.
 XX

Neisseria meningitidis.

WO9957280-A2.

11-NOV-1999.

30-APR-1999; 99WO-US09346.

01-MAY-1998; 98US-0083758.

31-JUL-1998; 98US-0094869.

02-SEP-1998; 98US-0098994.

02-SEP-1998; 98US-0099062.

09-OCT-1998; 98US-0103749.

09-OCT-1998; 98US-0103794.

25-FEB-1999; 99US-0121528.

(CHIR) CHIRON CORP.
 PA (GENO-) INST GENOMIC RES.

PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelein H, Venter JC;
XX
XX MPI; 2000-062150/05.
DR N-PSDB; AAZ54511.
XX
XX Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
XX
PS Claim 2; Page 1390; 1453pp; English.

AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
PCR primers used in the exemplification of the present invention. The
polypeptides, the polynucleotides, antibodies and compositions of
the invention can be used as vaccines, as diagnostic reagents, and as
immunogenic compositions. The polypeptides can be used in the
manufacture of medicaments for treating or preventing infection due to
Neisserial bacteria (e.g. meningitis and septicemia), to detect the
presence of Neisseria bacteria, or to raise antibodies. They may also
be used to screen for agonists or antagonists, which may themselves
have use as antibacterial agents. The polynucleotides of the invention
may also be used in gene therapy protocols.

Sequence 499 AA;

Query Match 95.5%; Score 2248; DB 21; Length 499;

Best Local Similarity 96.1%; Pred. No. 1.4e-181;
Matches 445; Conservative 5; Mismatches 13; Indels 0; Gaps 0;

QY 1 VFKKQYFALAAALCAALLAGCEKASFFGADKKEASFVERIHTKDDGSVSMILPFAQL 60
DB 1 VFKKQYFALAAALCAALLAGCEKASFFGADKKEASFVERIHTKDDGSVSMILPFAQL 60
QY 61 VOSEGPAAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVPMNPEIPOEADD 120
DB 61 VOSEGPAAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVPMNPEIPOEADD 120
QY 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIGSVQSDVALLKIDA 180
DB 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIGSVQSDVALLKIDA 180
QY 181 TEELPVVKIGNPKNLKGEWVAALGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTIVA 240
DB 181 TEELPVVKIGNPKNLKGEWVAALGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTIVA 240
QY 241 INFNGSGGFLPNLKQOVVINSQIYSRSGFMGISFAIPIDVAMNVAEQLNKTKVQRGQ 300
DB 241 INFNGSGGFLPNLKQOVVINSQIYSRSGFMGISFAIPIDVAMNVAEQLNKTKVQRGQ 300
QY 301 LGVITIGEVSVGLAOSFGDLKASGALIKLIPGSPABRAGLQAGDVIYSLDGEIRSSGDL 360
DB 301 LGVITIGEVSVGLAOSFGDLKASGALIKLIPGSPABRAGLQAGDVIYSLDGEIRSSGDL 360
QY 361 PVMVGAITPEKEVSLGYWRKGEETITIKALGNAAEHTGASSKTEBAPTEOOGSTFVYES 420
DB 361 PVMVGAITPEKEVSLGYWRKGEETITIKALGNAAEHTGASSKTEBAPTEOOGSTFVYES 420
QY 421 AGITLQTHDSSGKHLVVVRVSDAERAGLRHGDILLAVRASP 463
DB 421 AGITLQTHDSSGKHLVVVRVSDAERAGLRHGDILLAVRASP 463

RESULT 5
AAZ5750 standard; Protein: 499 AA.
AAY75750;
AAZ5750;

DT 21-MAR-2000 (first entry)
XX

DE Neisseria meningitidis ORF 986 protein sequence SEQ ID NO:2972.
XX
XX Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antigenic; diagnostic; immunogenic; infection; meningitis; septicemia;
KW antibacterial; gene therapy.
XX
XX
OS Neisseria meningitidis.
XX
XX
PN WO957280-A2.
XX
PD 11-NOV-1999.
XX
XX 30-APR-1999; 99WO-US09346.
PF

XX 01-MAY-1998; 98US-0083758.
XX 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
XX (CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
XX
XX
PI Fraser C, Galeotti C, Grandi G, Hickey E, Maignani V, Mora M;
PI Petersen J, Piza M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelein H, Venter JC;
XX
XX MPI; 2000-062150/05.
DR N-PSDB; AAZ54512.
XX

Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -

Claim 2; Page 1392-1393; 1453pp; English.

AAZ53015 to AAZ54536, AAZ54577 to AAZ54615, and AAY74253 to AAY75941
represent novel Neisseria meningitis and N. gonorrhoeae polynucleotides
and polypeptides. AAZ54537 to AAZ54576 and AAZ54616 to AAZ5473 represent
PCR primers used in the exemplification of the present invention. The
polypeptides, the polynucleotides, antibodies and compositions of
the invention can be used as vaccines, as diagnostic reagents, and as
immunogenic compositions. The polypeptides can be used in the
manufacture of medicaments for treating or preventing infection due to
Neisserial bacteria (e.g. meningitis and septicemia), to detect the
presence of Neisseria bacteria, or to raise antibodies. They may also
be used to screen for agonists or antagonists, which may themselves
have use as antibacterial agents. The polynucleotides of the invention
may also be used in gene therapy protocols.

Sequence 499 AA;

Query Match 95.4%; Score 2244; DB 21; Length 499;

Best Local Similarity 95.9%; Pred. No. 3.1e-181;
Matches 444; Conservative 5; Mismatches 14; Indels 0; Gaps 0;

QY 1 VFKKQYFALAAALCAALLAGCEKASFFGADKKEASFVERIHTKDDGSVSMILPFAQL 60
DB 1 VFKKQYFALAAALCAALLAGCEKASFFGADKKEASFVERIHTKDDGSVSMILPFAQL 60
QY 61 VOSEGPAAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVPMNPEIPOEADD 120
DB 61 VOSEGPAAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVPMNPEIPOEADD 120
QY 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIGSDVQSDVALLKIDA 180
DB 121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIGSDVQSDVALLKIDA 180
QY 181 TEELPVVKIGNPKNLKGEWVAALGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTIVA 240
DB 181 TEELPVVKIGNPKNLKGEWVAALGAPFGFDSVTAGIVSAKGRSLPNESYTPFIQTIVA 240

QY 241 INPNSGGPLFNLKGQVVGINSQIYSRSGGFMGISFALPIDVANVVAEQKNTGKVQRGQ 300
FT /note= "Encoded by CGT"
FT Misc-difference 463
Db 241 INPNSGGPLFNLKGQVVGINSQIYSRSGGFMGISFALPIDVANVVAEQKNTGKVQRGQ 300
FT /note= "Encoded by CAA"
FT Misc-difference 465.498
QY 301 LGVITQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIYLSLDGGEIRSSGDL 360
FT /note= "Encoded by AGGCAAGTCCCGTCGAA"
FT PN W0200012535-A2.
XX
QY 361 PVMGAIITPGKEVSLGVMRKGEITIKAKLGNAAEHTGASKTDEAPYTEOOSGTFSVES 420
FT /note= "Encoded by AGGCAAGTCCCGTCGAA"
Db 361 PVMGAIITPGKEVSLGVMRKGEITIKAKLGNAAEHTGASKTDEAPYTEOOSGTFSVES 420
QY 421 AGITLQHTDSSGKHLVVVRVSDAERAGLRHGEIILAVRAS 463
Db 421 AGITLQHTDSSGKHLVVVRVSDAERAGLRHGEIILAVRAS 463
RESULT 6
0413
AA70413 standard; Protein; 498 AA.
AC AA70413;
XX
XX 03-JUN-2000 (first entry)
XX
DE Neisseria meningitidis NMASP protein-2.
XX
XX NMASP, non-cytosolic; antibacterial; antiinflammatory; cytotoxic;
KM anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;
XX Neisserial infection; meningitidis; septicemia.
XX
XX OS Neisseria meningitidis.
FH Key Location/Qualifiers
FT Misc-difference 8 /note= "Encoded by TTC"
FT Misc-difference 17 /note= "Encoded by TTG"
FT Misc-difference 22 /note= "Encoded by GAA"
FT Misc-difference 42 /note= "Encoded by GAA"
FT Misc-difference 58 /note= "Encoded by GAA"
FT Misc-difference 84 /note= "Encoded by GCC"
FT Misc-difference 134 /note= "Encoded by GGC"
FT Misc-difference 145 /note= "Encoded by AAC"
FT Misc-difference 194 /note= "Encoded by GCC"
FT Misc-difference 218 /note= "Encoded by AAT"
FT Misc-difference 241 /note= "Encoded by GGCATC"
FT Misc-difference 340 /note= "Encoded by AGC"
FT Misc-difference 387 /note= "Encoded by AGG"
FT Misc-difference 396 /note= "Encoded by GGC"
FT Misc-difference 433 /note= "Encoded by ACC"
FT Misc-difference 451 /note= "Encoded by AAA"
FT Misc-difference 459 /note= "Encoded by CAC"
FT Misc-difference 460 /note= "Encoded by AGG"
FT Misc-difference 461 /note= "Encoded by GCA"
FT /note= "Encoded by AGT"

FT Misc-difference 463
FT /note= "Encoded by CGT"
FT Misc-difference 464
FT /note= "Encoded by CAA"
FT Misc-difference 465.498
FT /note= "Encoded by AGGCAAGTCCCGTCGAA"
XX
XX W0200012535-A2.
XX
XX 09-MAR-2000.
XX
XX 01-SEP-1999; 99W0-US19663.
XX
XX 01-SEP-1998; 98US-0098685.
XX
XX (ANTE-) ANTEK BIOLOGICS INC.
XX
XX Jackson WJ, Harris AM;
XX
XX WPI; 2000-256581/22.
XX
XX N-PSDB; AA251538.
XX
XX Neisseria meningitidis NMASP polypeptide, nucleotide sequences and
XX antibodies, useful in vaccines against infection
XX
XX Claim 5; Page 70-71; 75pp; English.
XX
XX The present sequence is the Neisseria meningitidis NMASP protein.
XX NMASP is a non-cytosolic protein, with antibacterial and
XX antiinflammatory activity. It shows sequence similarity to E. coli
XX Degr (HtrA) protein. NMASP proteins can be used as ligands to detect
XX antibodies elicited in response to N. meningitidis infections. Cytotoxic
XX anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA
XX may be used for diagnosis, therapy or prophylaxis of Neisserial
XX infections such as, bacterial meningitidis and septicemia.
XX
XX Sequence 498 AA;
SQ
Query Match 95.0%; Score 2234.5; DB 21; Length 498;
Best Local Similarity 95.9%; Pred. No. 1.9e-180;
Matches 444; Conservative 5; Mismatches 13; Indels 1; Gaps 1;
QY 1 VFKKYQYFALALCAALLAGCEKAGSFFGADKKEASFEVERIEHTKDDGSVMTLPDPAQL 60
Db 1 VFKKYQYFALALCAALLAGCEKAGSFFGADKKEASFEVERIEHTKDDGSVMTLPDPAQL 60
QY 61 VQSEGPVAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVNMPEIPOEABD 120
Db 61 VQSEGPVAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVNMPEIPOEABD 120
QY 121 GGLNFGSGFLISKNGYILTNTHVAVAGMSIVLNDKREYTAKLIGSDVQSVALLKIDA 180
Db 121 GGLNFGSGFLISKNGYILTNTHVAVAGMSIVLNDKREYTAKLIGSDVQSVALLKIDA 180
QY 181 TEELPVVKIKGNPKULKPGEMVAAICAPFGFDSVTAGIVSAKGRSLPESYTPFIQDVA 240
Db 181 TEELPVVKIKGNPKULKPGEMVAAICAPFGFDSVTAGIVSAKGRSLPESYTPFIQDVA 240
QY 241 INPNSGGPLFNLKGQVVGINSQIYSRSGGFMGISFALPIDVANVVAEQKNTGKVQRGQ 300
Db 241 INPNSGGPLFNLKGQVVGINSQIYSRSGGFMGISFALPIDVANVVAEQKNTGKVQRGQ 300
QY 301 LGVITQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIYLSLDGGEIRSSGDL 360
Db 301 LGVITQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIYLSLDGGEIRSSGDL 360
QY 361 PVMGAIITPGKEVSLGVMRKGEITIKAKLGNAAEHTGASKTDEAPYTEOOSGTFSVES 420
Db 361 PVMGAIITPGKEVSLGVMRKGEITIKAKLGNAAEHTGASKTDEAPYTEOOSGTFSVES 420
QY 421 AGITLQHTDSSGKHLVVVRVSDAERAGLRHGEIILAVRAS 463
Db 421 AGITLQHTDSSGKHLVVVRVSDAERAGLRHGEIILAVRAS 463

```
RESULT 7
AAV52994
ID AAV52994 standard; Protein: 499 AA.
XX
AC AAV52994;
XX
DT 21-FEB-2000 (first entry)
XX
DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.
XX
KW Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
KM antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
XX
XX invasive bacterial disease; antibacterial.
XX
OS Neisseria meningitidis.
XX
PN WO955872-A1.
XX
04-NOV-1999.
XX
20-APR-1999; 99WO-EP02765.
XX
23-APR-1998; 98GB-0008734.
XX
(SM1K ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
Ruelle J;
XX
WPI; 2000-052809/04.
XX
N-PSDB; AA233306.
XX
PT Novel polynucleotides and polypeptides from Neisseria meningitis used
XX
to prepare vaccines against bacterial infections -
XX
Claim 3; Page 77-78; 94pp; English.
XX
XX
XX The present sequence represents a BASB013 polypeptide isolated from
CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be
CC employed as research reagents and material for the discovery of
CC treatments and diagnostics for diseases, particularly human diseases.
CC They can be used for diagnosis of disease, staging of disease, or
CC determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and
CC for screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB013 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
CC purposes. The polypeptides can be used to produce antibodies. The
CC polypeptides can also be used in vaccine formulations, and to identify
CC agonists and antagonists. The polypeptides, antibodies, agonists and
CC antagonists (which are bacteriostatic) are used for the treatment and
CC prevention of diseases such as upper respiratory tract infection,
CC invasive bacterial diseases such as bacteraemia and meningitis, and for
CC the development and screening of antibacterial drugs. They are also used
CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins
CC on in-dwelling devices, or to extracellular proteins on wounds, and to
CC thus prevent tissue damage and/or block the normal progression of
CC pathogenesis in infections initiated other than by the implantation of
CC in-dwelling devices or by other surgical techniques.
XX
SQ Sequence 499 AA;
XX
XX
XX Query Match 93.6%; Score 2202; DB 21; Length 499;
XX Best Local Similarity 94.0%; Pred. No. 1,1e-177;
XX Matches 435; Conservative 10; Mismatches 18; Indels 0; Gaps 0;
XX
XX
XX 1 VFKKYVYFALAAACALLACCEKAGSFPGADKKEASFIERIEHKRDKGSVSMILPDPFQQL 60
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1 MKKKYIYLALALCAASLACGDKRAGSFGADKKEASFEVERIEHKRDKDSVSMILPDPFQQL 60
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
XX 61 VQSEGPVVNIQAPAPRTONGSGNAETDSDPLADSDPFYEFPRQLVPMPEIQEADAD 120
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
```

```
DB 61 VQSEGPVVNIQAPAPRTONGSGNAETDSDPLADSDPFYEFPRQLVPMPEIQEADAD 120
OY
DB 121 GGLNFGSGFIISKNGYIILNTHVAVAGMSIKYLLNDKREYTAKLIGSDVQSDVALLKIDA 180
DB 121 GGLNFGSGFIISKNGYIILNTHVAVAGMSIKYLLNDKREYTAKLIGSDVQSDVALLKIDA 180
OY
DB 181 TEELPVRKIGNPKNIKPEBWVAIIGAPFGFDSVTAGIVSAKGRSLIPNESYTFPIQTDVA 240
DB 181 TEELPVRKIGNPKNIKPEBWVAIIGAPFGFDSVTAGIVSAKGRSLIPNESYTFPIQTDVA 240
OY
DB 241 INPNSGGPLFNLKQOVGINSOIYRSRSGFENGISPAIPIDVAMNVAEOLKXNGXQORQ 300
DB 241 INPNSGGPLFNLKQOVGINSOIYRSRSGFENGISPAIPIDVAMNVAEOLKXNGXQORQ 300
OY
DB 301 LGVIIQEVSYGLAQSFGDLKASGALIAKILPSPPARAGLQAGDIYLSIDGGEIRSSGDL 360
DB 301 LGVIIQEVSYGLAQSFGDLKASGALIAKILPSPPARAGLQAGDIYLSIDGGEIRSSGDL 360
OY
DB 361 PVMVGAIITPGKEVSLGVWRKGEITTKAKLGNAEHTGASSTDEAPYTEQSGTFSVES 420
DB 361 PVMVGAIITPGKEVSLGVWRKGEITTKAKLGNAEHTGASSTDEAPYTEQSGTFSVES 420
OY
DB 421 AGITLQHTDSSGKHLVVRVSDAERAGLRGCDETLAVRAGP 463
DB 421 AGITLQHTDSSGKHLVVRVSDAERAGLRGCDETLAVRAGP 463
OY
DB 421 AGITLQHTDSSGKHLVVRVSDAERAGLRGCDETLAVRAGP 463
DB 421 AGITLQHTDSSGKHLVVRVSDAERAGLRGCDETLAVRAGP 463
XX
XX RESULT 8
XX AAV52993
XX ID AAV52993 standard; Protein: 499 AA.
XX
XX AC AAV52993;
XX
XX DT 21-FEB-2000 (first entry)
XX
XX DE Neisseria meningitidis strain ATCC 13090 BASB013 protein sequence.
XX
XX
XX Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;
XX KM antibiotic; upper respiratory tract infection; bacteraemia; meningitis;
XX KM invasive bacterial disease; antibacterial.
XX
XX OS Neisseria meningitidis.
XX
XX PN WO955872-A1.
XX
XX 04-NOV-1999.
XX
XX 20-APR-1999; 99WO-EP02765.
XX
XX 23-APR-1998; 98GB-0008734.
XX
XX (SM1K ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Ruelle J;
XX
XX WPI; 2000-052809/04.
XX
XX N-PSDB; AA233305.
XX
XX PT Novel polynucleotides and polypeptides from Neisseria meningitis used
XX to prepare vaccines against bacterial infections -
XX
XX Claim 5; Page 74-76; 94pp; English.
XX
XX
XX The present sequence represents a BASB013 polypeptide isolated from
CC Neisseria meningitidis. BASB013 polynucleotides and polypeptides may be
CC employed as research reagents and material for the discovery of
CC treatments and diagnostics for diseases, particularly human diseases.
CC They can be used for diagnosis of disease, staging of disease, or
CC determining response of an infectious organism to drugs. The
CC polynucleotides may be used as a source for hybridisation probes, and
CC for screening of genetic mutations, serotype, organism or strain
CC identification, identification of mutation in BASB013 sequences, and as
CC components of arrays which are useful for diagnostic and prognostic
```

CC purposes. The polypeptides can be used to produce antibodies. The
 CC polypeptides can also be used in vaccine formulations, and to identify
 CC agonists and antagonists. The polypeptides, antibodies, agonists and
 CC antagonists (which are bacteriostatic) are used for the treatment and
 CC prevention of diseases such as upper respiratory tract infection,
 CC invasive bacterial diseases such as bacteraemia and meningitis, and for
 CC the development and screening of antibacterial drugs. They are also used
 CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins
 CC on in-dwelling devices, or to extracellular proteins on wounds, and to
 CC thus prevent tissue damage and/or block the normal progression of
 CC pathogenesis in infections initiated other than by the implantation of
 CC in-dwelling devices or by other surgical techniques.

XX Sequence 499 AA:

Query Match 93.4%; Score 2197; DB 21; Length 499;
 Best Local Similarity 93.7%; Pred. No. 2.9e-177;
 Matches 434; Conservative 10; Mismatches 19; Indels 0; Gaps 0;

```

1 VFKKYQYALALCALLAGCEKAGSPFGADKKEASFVERIHTKDDGSVSMLLPDPQL 60
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1 MFKKYQYALALCAASLAGCDKAGSPFGADKKEASFVERIHTKDDGSVSMLLPDPQL 60
61 VQSEGPVVNIQAAAPARTONGSNAETDSDPLADSDPFYEFKRLVPMPEIPOBEAD 120
61 VQSEGPVVNIQAAAPARTONGSNAETDSDPLADSDPFYEFKRLVPMPEIPOBEAD 120
121 GGLNFGSGFIISKNGYILTNTHVVAAGMSIKVLLNDKREYTKLIGSDVQSVALLKIDA 180
121 GGLNFGSGFIISKNGYILTNTHVVTGMGSIKVLLNDKREYTKLIGSDVQSVALLKIDA 180
181 TRELTVYKIGNKXNKPGEVVAATGAPGPDNSVTAGIYSAKGRSLPNSYPIQTDVA 240
181 TRELTVYKIGNKXNKPGEVVAATGAPGPDNSVTAGIYSAKGRSLPNSYPIQTDVA 240
191 TEELPVYKIGNKPDLPGEVVAATGAPGPDNSVTAGIYSAKGRSLPNSYPIQTDVA 240
241 INPGNSGGLFLNKLKQVVGINSQIYSRSGFMGISPAIPIDVAMVNAEOLKNTGKVQRCQ 300
241 INPGNSGGLFLNKLKQVVGINSQIYSRSGFMGISPAIPIDVAMVNAEOLKNTGKVQRCQ 300
301 LGVITIOEVSYGLAQSGFLDKASGALIAKILPGSPARRAGIQADITLSLDGGEIRSSGDL 360
301 LGVITIOEVSYGLAQSGFLDKAGGALIAKILPGSPARRAGIQADITLSLDGGEIRSSGDL 360
361 PVMVGAITGKEKSLGVWRKGEKITIKAKLGNAAEHTGASSTKDEAPYTEEQSGTSVES 420
361 PVMVGAITGKEKSLGVWRKGEKITIKAKLGNAAEHTGASSTKDEAPYTEEQSGTSVES 420
421 AGITLQTHTDSGSKHLVVVRVSDAERAGLRHGDDEILAVRASP 463
421 AGITLQTHTDSGSKHLVVVRVSDAERAGLRHGDDEILAVRASP 463
421 AGITLQTHTDSGSKRLVVVRVSGAERAGLRGDDEILAVGQVP 463

```

RESULT 9
 AAY70414
 ID AAY70414 standard; Protein; 475 AA.

AC AAY70414;
 XX
 XX
 DT 03-JUL-2000 (first entry)

DE Neisseria meningitidis NMASP protein-3.

XX NMASP: non-cytosolic; antibacterial; antiinflammatory; cytotoxic;
 KW anti-NMASP antibody; vaccine; diagnosis; therapy; prophylaxis;
 XX Neisserial infection; meningitidis; septicemia.

XX Neisseria meningitidis.

XX Key Location/Qualifiers

FT Misc-difference 19 /note= "Encoded by GAA"

FT Misc-difference 35 /note= "Encoded by GCC"

```

FT Misc-difference 61 /note= "Encoded by GGC"
FT Misc-difference 111 /note= "Encoded by AAC"
FT Misc-difference 122 /note= "Encoded by GCC"
FT Misc-difference 171 /note= "Encoded by GCA"
FT Misc-difference 171 /note= "Encoded by AAT"
FT Misc-difference 194 /note= "Encoded by GGCATC"
FT Misc-difference 298 /note= "Encoded by AGC"
FT Misc-difference 317 /note= "Encoded by AGC"
FT Misc-difference 364 /note= "Encoded by AGC"
FT Misc-difference 373 /note= "Encoded by GCC"
FT Misc-difference 410 /note= "Encoded by ACC"
FT Misc-difference 428 /note= "Encoded by AAA"
FT Misc-difference 436 /note= "Encoded by CAC"
FT Misc-difference 437 /note= "Encoded by AGG"
FT Misc-difference 438 /note= "Encoded by GCA"
FT Misc-difference 440 /note= "Encoded by AGT"
FT Misc-difference 440 /note= "Encoded by CCGCA"

```

PN WO200012535-A2.

PD 09-MAR-2000.

PF 01-SEP-1999; 99WO-US19663.

PR 01-SEP-1999; 98US-0098685.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Jackson WJ, Harris AM;

DR WPI; 2000-256581/22.

XX N-PSDB; AA251539.

PT Neisseria meningitidis NMASP polypeptide, nucleotide sequences and
 antibodies, useful in vaccines against infection -

PS Claim 5; Page 71-73; 75pp; English.

XX The present sequence is the Neisseria meningitidis NMASP protein.
 CC NMASP is a non-cytosolic protein, with antibacterial and
 CC antiinflammatory activity. It shows sequence similarity to E. coli
 CC Degr (HcrA) protein. NMASP proteins can be used as ligands to detect
 CC antibodies elicited in response to N. meningitidis infections. Cytotoxic
 CC anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA
 CC may be used for diagnosis, therapy or prophylaxis of Neisserial
 CC infections such as, bacterial meningitidis and septicemia.

XX Sequence 475 AA:

Query Match 90.5%; Score 2130.5; DB 21; Length 475;
 Best Local Similarity 96.4%; Pred. No. 1.2e-171;
 Matches 424; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

QY 24 AGSPFGADKKEASFVERIHTKDDGSVSMLLPDPQLVQSEGPVVNIQAAAPARTONGS 83

DB 1 AGSPFGADKKEASFVERIHTKDDGSVSMLLPDPQLVQSEGPVVNIQAAAPARTONGS 60

QY 84 GNAETDSDPLADSDPFYEFKRLVPMPEIPOBEADGILNFGSGFIISKNGYILTNTHV 143
 DB 61 SNAETDSDPLADSDPFYEFKRLVPMPEIPOBEADGILNFGSGFIISKNGYILTNTHV 120

QY 144 VAGMGSIKVLNDKREYTAKLIGSDVQSDVALLKIDATEELPVVKIGNPKLKEGEWAA 203
 DB 121 VTGMSGIKVLNDKREYTAKLIGSDVQSDVALLKIDATEELPVVKIGNPKLKEGEWAA 180
 QY 204 ICAPFGFDSVTAGIYSAKGRSLPNESYTPFIQTVDVAINPNSGGLPVLNKGQVVGINSQ 263
 DB 181 IGAPFGFDSVTAGIYSAKGRSLPNESYTPFIQTVDVAINPNSGGLPVLNKGQVVGINSQ 239
 QY 264 IYRSRGFGMGISPAIPFDVAMNVAEQLKNTGKVRGGLGVIIQSVSGLAOSFGLDKASG 323
 DB 240 IYRSRGFGMGISPAIPFDVAMNVAEQLKNTGKVRGGLGVIIQSVSGLAOSFGLDKASG 299
 QY 324 ALIAXILPGSPARAGLQAGDIVLSLDGEIRSSGDLPMVVGAIITPKKEVSLGYWRKGE 383
 DB 300 ALIAXILPGSPARAGLQAGDIVLSLDGEIRSSGDLPMVVGAIITPKKEVSLGYWRKGE 359
 QY 384 ITIKAKLGNAAEHTGASSKTDEAPYTEQSGTSVESAGITLQTHTDSGKHLVVRVSD 443
 DB 360 ITIKAKLGNAAEHTGASSKTDEAPYTEQSGTSVESAGITLQTHTDSGKHLVVRVSD 419
 444 AABRAGLRHGDDELAVASP 463
 DB 420 AABRAGLRHGDDELAVASP 439

RESULT 10

ID AAY83151
 AC AAY83151 standard; Protein: 414 AA.

DT 24-JUL-2000 (first entry)

DE NGSP polypeptide of *Neisseria gonorrhoeae*.

KW NGSP, polypeptide; peptide; vaccine; immune response; antibody;
 cellular matrix; adherence domain; ligand; detection; diagnosis;
 screening; probe; primer; prophylaxis; therapy.

OS *Neisseria gonorrhoeae*.

PN WO200012133-A1.

PD 09-MAR-2000.

PF 01-SEP-1999; 99WO-US20070.

PY 01-SEP-1998; 98US-0098685.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Jackson WJ, Harris AM;

DR WPI; 2000-237782/20.

DR N-PSDB; AA293415.

PT Non-cytosolic NGSP polypeptide and polynucleotide sequence from
 PT *Neisseria* useful for diagnosis, prevention or treatment of *Neisseria*
 PT infections

PS Claim 5; Page 63-64; 68pp; English.

CC The NGSP polypeptide of *N. gonorrhoeae* has conserved Arg-Gly-Asp and
 CC Arg-Gly-Asn motifs near the C-terminus which function as adherence
 CC domains for extracellular matrix proteins. Using the NGSP polypeptide
 CC as a vaccine produces antibodies which inhibit binding of *N.*
 CC *gonorrhoeae* to the host's cellular matrix reducing attachment and/or
 CC subsequent invasion. The NGSP polypeptide and its peptide fragments
 CC can be used to immunize an animal and produce an immune response.
 CC They can also be used as ligands to detect antibodies elicited in
 CC response to *Neisseria* infections and also as antigens or immunogens
 CC for inducing *Neisseria*-specific antibodies which are useful in

CC immunosays to detect *Neisseria* in biological specimens. Nucleotides
 CC encoding NGSP or its fragments can be used as probes to identify
 CC *Neisseria* in biological specimens by hybridization or polymerase
 CC chain reaction amplification. The NGSP polypeptide can also be used
 CC in screening assays to identify agents and compounds which useful as
 CC diagnostic, prophylactic or therapeutic agents against *Neisseria*
 CC infection.

SQ Sequence 414 AA;

Query Match 88.9%; Score 2092; DB 21; Length 414;

Best Local Similarity 100.0%; Pred. No. 1,7e-168;

Matches 414; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 52 MLPPFAQLVQSEGAANNVIAAPAPRTONGSGNAETBDDPLADSDPFYEFKRLVPPNP 111
 DB 1 MLPPFAQLVQSEGAANNVIAAPAPRTONGSGNAETBDDPLADSDPFYEFKRLVPPNP 60
 QY 112 EIPQEADDDGGLNFGSGFIIISKNGYIINTHVAGMGSIKVLNDKREYTAKLIGSDVQS 171
 DB 61 EIPQEADDDGGLNFGSGFIIISKNGYIINTHVAGMGSIKVLNDKREYTAKLIGSDVQS 120
 QY 172 DVALLKIDATEELPVVKIGNPKLKEGEWAAIGAPFGFDSVTAGIYSAKGRSLPNESY 231
 DB 121 DVALLKIDATEELPVVKIGNPKLKEGEWAAIGAPFGFDSVTAGIYSAKGRSLPNESY 180
 QY 232 TPIQTDVAINPNSGGLPVLNKGQVVGINSQIYRSRGFGMGISPAIPFDVAMNVAEQLK 291
 DB 181 TPIQTDVAINPNSGGLPVLNKGQVVGINSQIYRSRGFGMGISPAIPFDVAMNVAEQLK 240
 QY 292 NTGKVRGQLGVIIQSVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIVLSLDG 351
 DB 241 NTGKVRGQLGVIIQSVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIVLSLDG 300
 QY 352 GEIRSSGGLPMVVGAIITPKKEVSLGYWRKGEITTIKAKLGNAAEHTGASSKTDEAPYTEQ 411
 DB 301 GEIRSSGGLPMVVGAIITPKKEVSLGYWRKGEITTIKAKLGNAAEHTGASSKTDEAPYTEQ 360
 QY 412 QSGTFSVESAGITLQTHTDSGKHLVVRVSDAABRAGLRHGDDELAVASP 465
 DB 361 QSGTFSVESAGITLQTHTDSGKHLVVRVSDAABRAGLRHGDDELAVASP 414

RESULT 11

ID AAY70409
 AC AAY70409 standard; Protein: 448 AA.

AC AAY70409;

DT 03-JUL-2000 (first entry)

DE *Neisseria meningitidis* NMA SP protein-1.

KW NMA SP, non-cytosolic; antibacterial; antiinflammatory; cytotoxic;

KW anti-NMA SP antibody; vaccine; diagnosis; therapy; prophylaxis;

KW *Neisseria* infection; meningitidis; septicemia.

OS *Neisseria meningitidis*.

FT Key Location/Qualifiers

FT Misc-difference 167

FT WO200012535-A2.

PD 09-MAR-2000.

PF 01-SEP-1999; 99WO-US19663.

PY 01-SEP-1998; 98US-0098685.

PA (ANTE-) ANTEX BIOLOGICS INC.

PI Jackson WJ, Harris AM;
 XX WPI; 2000-256581/22.
 DR N-PSDB; AA251533.
 XX
 PT Neisseria meningitidis NMASP polypeptide, nucleotide sequences and
 PT antibodies, useful in vaccines against infection
 XX
 PS Claim 5; Page-; 75pp; English.
 XX
 CC The present sequence is the Neisseria meningitidis NMASP protein.
 CC NMASP is a non-cytosolic protein, with antibacterial and
 CC antiinflammatory activity. It shows sequence similarity to E. coli
 CC DspA (HtrA) protein. NMASP proteins can be used as ligands to detect
 CC antibodies elicited in response to N. meningitidis infections. Cytotoxic
 CC anti-NMASP antibodies can be used as vaccines. NMASP proteins and DNA
 CC may be used for diagnosis, therapy or prophylaxis of Neisserial
 CC infections such as, bacterial meningitidis and septicemia.
 CC Note: The protein represented in SEQ ID NO:2 of the specification is
 CC erroneous. The present sequence is the decoded version of the nucleotide
 CC represented in AA251533.

SO Sequence 448 AA;

Query Match 85.1%; Score 2002; DB 21; Length 448;

Best Local Similarity 96.4%; Pred. No. 8e-16i; Mismatches 12; Indels 0; Gaps 0;

Matches 397; Conservative 3; Mismatches 12; Indels 0; Gaps 0;

52 MLDPFQVLSVSEGAIVNIQAAPARTONGSNAETDSDPLADSPFFEFKRLVPMNP 111

1 MLDPFQVLSVSEGAIVNIQAAPARTONGSNAETDSDPLADSPFFEFKRLVPMNP 60

112 EIPOEADGGINFGSGFIISKNGYILTNTHVVAAGMSIKVLNDRREYAKLIGSDVQS 171

61 EIPOEADGGINFGSGFIISKNGYILTNTHVVAAGMSIKVLNDRREYAKLIGSDVQS 120

172 DVALLKIDATEELPVYKIGNPNKLRGEWVAAGAPFGEDNSVTAGIVSAKGRSLPNEST 231

121 DVALLKIDATEELPVYKIGNPNKLRGEWVAAGAPFGEDNSVTAGIVSAKGRSLPNEST 180

232 TPFIOTDVAINPNSGGPFLNKGQVVGINSQIYSSGGFMGISPAIPDIVANVAEQLK 291

181 TPFIOTDVAINPNSGGPFLNKGQVVGINSQIYSSGGFMGISPAIPDIVANVAEQLK 240

292 NTGKVGQGLGVYIIOEVSYGLAQSGFLDRAKASGALIKILPGSPAERAGLQAGIVVSLDG 351

241 NTGKVGQGLGVYIIOEVSYGLAQSGFLDRAKASGALIKILPGSPAERAGLQAGIVVSLDG 300

352 GEIRSSGDLPVWVGAIITPGEKESVSLGWRKGEETITIRAKLGNAAEHTGASSKTDEAPYTEQ 411

301 GEIRSSGDLPVWVGAIITPGEKESVSLGWRKGEETITIRAKLGNAAEHTGASSKTDEAPYTEQ 360

412 QSGTFSVSAGITLQTHDSSGKHLVYVRSVDAEPAQLRHGDELLAVRASP 463

361 QSGTFSVSAGITLQTHDSSGKHLVYVRSVDAEPAQLRHGDELLAVRASP 412

RESULT 12

AA252996

ID AA252996 standard; Protein; 370 AA.

AC AA252996;

DT 21-FEB-2000 (first entry)

DE Neisseria meningitidis BASB013-C protein sequence.

XX Neisseria meningitidis; BASB013; diagnosis; infection; vaccine;

XX antibiotic; upper respiratory tract infection; bacteraemia; meningitis;

XX invasive bacterial disease; antibacterial.

OS Neisseria meningitidis.

PN WO955872-A1.
 XX
 XX 04-NOV-1999.
 XX
 PF 20-APR-1999; 99MO-EP02765.
 XX
 PR 23-APR-1998; 98GB-0008734.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Ruelle J;
 XX
 DR WPI; 2000-052809/04.
 DR N-PSDB; AA233308.
 XX
 PT Novel polynucleotides and polypeptides from Neisseria meningitidis used
 PT to prepare vaccines against bacterial infections
 XX
 PS Example 2; Page 82-83; 94pp; English.
 XX
 CC The present sequence is a conserved BASB013-C polypeptide isolated
 CC from Neisseria meningitidis. BASB013 polynucleotides and polypeptides
 CC may be employed as research reagents and material for the discovery of
 CC treatments and diagnostics for diseases, particularly human diseases.
 CC They can be used for diagnosis of disease, staging of disease, or
 CC determining response of an infectious organism to drugs. The
 CC polynucleotides may be used as a source for hybridisation probes, and
 CC for screening of genetic mutations, serotype, organism or strain
 CC identification, identification of mutation in BASB013 sequences, and as
 CC components of arrays which are useful for diagnostic and prognostic
 CC purposes. The polypeptides can be used to produce antibodies. The
 CC polypeptides can also be used in vaccine formulations, and to identify
 CC antagonists (which are bacteriostatic) are used for the treatment and
 CC prevention of diseases such as upper respiratory tract infection,
 CC invasive bacterial diseases such as bacteraemia and meningitis, and for
 CC the development and screening of antibacterial drugs. They are also used
 CC in the prevention of adhesion of bacteria to eukaryotic matrix proteins
 CC on in-dwelling devices, or to extracellular proteins on wounds, and to
 CC thus prevent tissue damage and/or block the normal progression of
 CC pathogenesis in infections initiated other than by the implantation of
 CC in-dwelling devices or by other surgical techniques.

SO Sequence 370 AA;

Query Match 77.9%; Score 1832; DB 21; Length 370;

Best Local Similarity 97.0%; Pred. No. 1.5e-146; Mismatches 5; Indels 0; Gaps 0;

Matches 359; Conservative 6; Mismatches 5; Indels 0; Gaps 0;

1 VFKKYGFALAAALCAALLAGCEKASFFGADKEASPFYRIEHTKDDGSVSMILDPFAQL 60

1 MFKKYQYLALAAALCAALLAGCEKASFFGADKEASPFYRIEHTKDDGSVSMILDPFAQL 60

61 VQSEGAIVNIQAAPARTONGSNAETDSDPLADSPFFEFKRLVPMNPPIQBEADD 120

61 VQSEGAIVNIQAAPARTONGSNAETDSDPLADSPFFEFKRLVPMNPPIQBEADD 120

121 GGLNFGSGFIISKNGYILTNTHVVAAGMSIKVLNDRREYAKLIGSDVQSVALLKIDA 180

121 GGLNFGSGFIISKNGYILTNTHVVAAGMSIKVLNDRREYAKLIGSDVQSVALLKIDA 180

181 TEELPVYKIGNPNKLRGEWVAAGAPFGEDNSVTAGIVSAKGRSLPNESTYPTFIOTDVA 240

181 TEELPVYKIGNPNKLRGEWVAAGAPFGEDNSVTAGIVSAKGRSLPNESTYPTFIOTDVA 240

241 INPNSGGPFLNKGQVVGINSQIYSSGGFMGISPAIPDIVANVAEQLKNTGKVGQRCQ 300

241 INPNSGGPFLNKGQVVGINSQIYSSGGFMGISPAIPDIVANVAEQLKNTGKVGQRCQ 300

301 LGVYIIOEVSYGLAQSGFLDRAKASGALIKILPGSPAERAGLQAGIVVSLDGIRISSGDL 360

301 LGVYIIOEVSYGLAQSGFLDRAKASGALIKILPGSPAERAGLQAGIVVSLDGIRISSGDL 360

```
OY 361 PWMGAIPTG 370
DB 361 PWMGAIPTG 370
RESULT 13
ID AAY29294 standard; Protein: 460 AA.
AC AAY29294;
XX 25-OCT-1999 (first entry)
DE Protein encoded by the PA14 degt gene.
XX Human pathogen; virulence polypeptide; virulence factor;
KW pathogenic infection; Pseudomonas aeruginosa infection.
OS Pseudomonas aeruginosa.
MO9927129-A1.
PD 03-JUN-1999.
PF 25-NOV-1998; 98WO-US25247.
PR 25-NOV-1997; 97US-0066517.
PA (GENO ) GEN HOSPITAL CORP.
P1 Ausubel F, Cao H, Drenkard E, Goodman HM, Mahajan-Miklos S;
PI Rahne LG, Tan M, Tsongalis J;
DR WPI: 1999-357851/30.
XX Virulence factors useful in developing disease treatments
PS Dieckhouse; Fig 28; 228pp; English.
XX The present sequence represents a Pseudomonas aeruginosa polypeptide
CC sequence. P. aeruginosa is an opportunistic human pathogen present in
CC soil water and plants. The specification describes virulence polypeptides
CC and nucleic acid sequence encoding such polypeptides. These sequences
CC can be used to identify a compound which is capable of decreasing the
CC expression of a pathogenic virulence factor. Compounds that inhibit
CC the expression or activity of virulence factor polypeptides can be
CC used to treat pathogenic infections, especially where the infection
CC is a P. aeruginosa infection.
CC note: the sequences given in the specification were poorly legible, and
CC in some instances assumptions were made as to the identity of the
CC residue; it is therefore possible that the sequence given below is
CC not entirely correct.
XX Sequence 460 AA;
SQ
Query Match 40.9%; Score 962.5; DB 20; Length 460;
Best Local Similarity 51.8%; Pred. No. 7.9e-73;
Matches 217; Conservative 56; Mismatches 113; Indels 33; Gaps 11;
OY 54 LPDPAQVOSGEPVNI---QAAPARTONGSGNAETSDPLADSDP-FYEFKRLVFN 109
DB 27 LPDFPLVQASPAVNVNISTROKLP---DRAMARQGLSIPDLGLPMPFRDPLERTIPQ 82
OY 110 MEIIP---OEADDDGLNPGSGFLLSKNGYILNTHVAVAGSGIKVLNDRKREYAKLIG 166
DB 83 VPRMPPRGQGRQEA---SLGGGFTISMDGYILTNHVVADDELIVRLSDSEHRAKLVG 138
OY 167 SDVQSDVALLKIDATEELPVVKIGNPKNLKPGEVVAIGAIFGPDNSVTAGIVSAKGRSL 226
DB 139 ADPRSDVAVLKIEA-KNLPTLKGDGSMKLGKGEVLAIGSPFGDHTTAAIVSAKGRSL 197
OY 227 PNESTPTPIQTDVAINPGNSGGLFNLKGVGVGINSQIYRSRSGFMGISFAIPIDVAMNV 286
```

```
DB 198 PMESVYPIQTDVAINPGNSGGLPLNLEGEVGVINSQIFTRSGFMGISFAIF
OY 287 AEQLKNTGKTVORQGLVITIOEVSYGLASFGLDKASGALIKLTPGSPARAGLOAGU..
DB 258 ADQKKRAGKVSRLGVVIOEVNKLAESEFGLDPSGHLVAVQVDEGPPAKGGLQVGDVI 317
OY 347 LSLDGEIRSSGDLPVWVGAITTPGKEVSLGYWRKGEIITIRAKLGNAAEHTGASSKTDE- 405
DB 318 LSLNGOSINESADLPHLVGMKKPGDKINLDVITRMGQRKSLMAVGNLPD-----DDBE 370
OY 406 -----APYTEQSGSTVESAGITLQTH--TDSSGKHLVYVRVSDA-ARAGLRHGDVI 456
DB 371 IASGAPGARSSRWLGVTVAADLTAEQRKSLDIQG-GVIVKEVDGPAAVIGLRPGDVI 428
RESULT 14
ID AAR77434 standard; Protein: 503 AA.
AC AAR77434;
XX 19-AUG-1996 (first entry)
DE Heat shock protein of Rochalimaea henselae.
XX Rochalimaea henselae; cat scratch disease; bacillary angiomatosis;
KW CSD; infection; antigen; antibody; vaccine.
OS Rochalimaea henselae.
XX W09531549-A1.
XX 23-NOV-1995.
PF 18-MAY-1995; 95WO-US06211.
PR 16-SEP-1994; 94US-0307279.
PR 18-MAY-1994; 94US-0245294.
XX (USSH ) US DEPT HEALTH & HUMAN SERVICES.
PA Anderson BE, Regnery RL;
PI
PT WPI: 1996-010935/01.
DR N-PSDB; AAT04403.
XX Nucleic acids of Rochalimaea henselae and R. quintana - methods which
PT enable the identification of R.henselae, which is a causative agent
PT of both cat scratch disease and bacillary angiomatosis
XX Claim 12; Page 84-86; 100pp; English.
XX The nucleic acids (AAT04402, AAT04403), fragments and antibodies
CC binding to the encoded proteins (AAR77433, AAR77434), may be used in the
CC diagnosis and detection of cat scratch disease (CSD) and bacillary
CC angiomatosis caused by R. henselae. The proteins or fragments of
CC them, may be used in vaccines to protect against R. henselae
CC infection.
XX Sequence 503 AA;
SQ
Query Match 31.1%; Score 731.5; DB 17; Length 503;
Best Local Similarity 39.5%; Pred. No. 3.2e-53;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;
OY 8 FALALCAALLACCEKASFFGADKKEASFVERIETHTKDDGSVSMLLPDAQVOSGEP 67
DB 15 FSAALRTALFFSGC---GSLMTTKAHANSV-----FSSIMQCGGFADIVGQKPA 62
OY 68 VVNIQAAPARTON---GSGNAETSDPLADSDP---FYEFKRLVNMMEIIPDEAD 120
DB 63 VVSQVQSNKKKEWFEFSDFTPGDLPDQHPLKKEFFDFYNRDKPSNXL-QRSRL 121
```

QY 121 GGLNGSGFIIISKNGYILNTHTVAVAGMSIKVLLNDKREYTAKLIGSDVQSVALLKIDA 180
 Db 122 RPIAGSGSFIISSDGYIVTNNHVIEGADSVRVNLEGTSGKEESLPAYI 136
 QY 181 TEELPVVXIGNPKNLKPGEWVAAIGAPFPGDNSVTAGIVSAKGRSLPNESYTFIQTDA 240
 Db 182 KRKFSYVPGDGSKLKRVGDVVAIGNPFGLGTVTIGIVSARGRDVGITGVYDDFIQIDAA 241
 QY 241 INPNSGGPFLFNKGOVVGINSQIYRSRGFMGISFPAIPIDVAMNVAEQLKNTGKYOQ 300
 Db 242 VNRGNSGPTFIDLNGKVVGVNTAIFSPSGANGIAIPAATANEVVOQLIEKGLVQRGW 301
 QY 301 LGVITIGFVSYGIAQSFGLDKASGALIAKILPGSPARAGIQAGDIYLSLDGGRSSGDL 360
 Db 302 LGVQIQPVTKETISDIGHKEAKGALITDPLKG-PAKAGIKAGDVIIISVGEKINDVRL 360
 QY 361 PVMGAIITPGKEVSLGVWRKGEETIKAKLGNAEHTGASSKTDEAPYR-EQSGSFPSV 419
 Db 361 AKRIANMSPGEVITIGVWKSKEENIKVKLDSMPED--ENMKDGSKYSNEHNSDETLE 417
 QY 420 SAGITLQHTDSSGKHLVVRV---SDAERAGLRHGDHILAV 459
 Db 418 DYGLIVAPSDDLG--LVVTVDVDPDSDADK-GIRGDIYIVY 457

RESULT 15

AAG78605
 ID AAG78605 standard; Protein; 474 AA.

AC AAG78605;

DT 20-NOV-2001 (first entry)

DE Lawsonia intracellularis protein SEQ ID NO: 7.

XX HtrA; PonA; HycC; YefW; ABC1; Omp100; Lawsonia intracellularis infection;
 KM vaccine.

OS Lawsonia intracellularis.

PN JP2001169787-A.

PD 26-JUN-2001.

PF 20-OCT-2000; 2000JP-0320736.

PR 22-OCT-1999; 99US-0160922.

(PFIZ) PFIZER PROD INC.

WPI; 2001-592540/67.

PT Lawsonia intracellularis polynucleotide and encoded protein, used to
 prevent Lawsonia intracellularis infection -

PS Claim 12; Page 47-49; 67pp; Japanese.

CC The present invention provides isolated polynucleotides encoding HtrA,
 CC PonA, HycC, YefW, ABC1 or Omp100 protein of Lawsonia
 CC intracellularis. The sequences can be used in vaccines for the prevention
 CC of Lawsonia intracellularis infection. The present sequence is a protein
 CC of the invention.

XX Sequence 474 AA;

Query Match 30.0%; Score 705; DB 22; Length 474;

Best Local Similarity 39.3%; Pred. No. 5,1e-51;
 Matches 169; Conservative 67; Mismatches 152; Indels 42; Gaps 11;

QY 54 LPDFAQDVOSGPAVWNIOA-APAFRTONGSGNAETDSDPLADSDPFYE-FKRLVPPNMP 111
 Db 28 LPNFVPLVWASKAVVNISTEKIIPR-----GRTEFPMEMFRGLPGRFRRFQGFEPKGP 82

QY 112 --EIPOEADDDGILNFGSGFIIISKNGYILNTHTVAVAGMSIKVLL---NDKREYTAKL 165
 Db 83 DSQIHKQR-----SLGTGFIISDGYIVTNNHVIEGADSVRVNLEGTSGKEESLPAYI 136
 QY 166 GSDVQSVALLKIDATEELPVVKIGNPKNLKPGEWVAAIGAPFPGDNSVTAGIVSAKGRS 225
 Db 137 GRDEETDLALLKVKSKDSLPLYLIFGNSDITMEVGEMVLAIGNPFGLGHTVIGILSAKGRD 196
 QY 226 LPNESYTFPIQTDAVAINPNSGGPFLFNKGOVVGINSQIYRSRGFMGISFPAIPIDVAMN 285
 Db 197 IHAGPFUNFIQTDAVINPNSGGPFLINSGOVVGINTAIMA-SG--QIGFAPISSMADR 253
 QY 286 VAEQLKNTGKIVORQGLVITIGFVSYGIAQSFGLDKASGALIAKILPGSPARAGIQAGDI 345
 Db 254 IIEQLKNTKKIVSGWIGVITIQDVTNTAKALGLSQAKGALVGSVPPGDPADKAGIKVGD 313
 QY 346 VLSLDGGRSSGDLPMVGAIITPGKEVSLGVWRKGEETIKAKLGNAEHTGASSKTDE 405
 Db 314 VTQADGKQIDSSASLLKAIATKPPFSVVKLKWNRDGSKDISITLGERKTTSSQKQSSPE 373
 QY 406 A-----PYTEOQSGTFESVSAGITLQHTDSSGKHLVVRVSDAERAGLRHGD 455
 Db 374 SLPGALGLSVRPLTQESKSFVVK-LGI-----GLVVSVEPPNPAEAGIRBODI 423
 QY 456 ILAVRASPRQ 465
 Db 424 ILSANLKPLQ 433

Search completed: July 11, 2003, 10:31:20
 Job time : 75 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:30:07 ; Search time 27 Seconds
(without alignments)
506.728 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKKYQIPALALCALIAG.....ERAGLRHDEILAVASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

1 number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	990.5	42.1	460	4	US-09-199-637A-132
2	731.5	31.1	503	1	US-08-245-294-8
3	731.5	31.1	503	1	US-08-474-499-8
4	731.5	31.1	503	1	US-08-307-279A-8
5	731.5	31.1	503	4	US-09-525-310-8
6	731.5	31.1	503	5	PCT-US95-06211-8
7	646	27.5	463	1	US-08-485-569-2
8	646	27.5	463	1	US-08-480-993-2
9	646	27.5	463	2	US-07-903-079B-2
10	633.5	26.9	475	1	US-08-350-741-2
11	633.5	26.9	475	1	US-08-463-875A-2
12	632	26.9	463	1	US-08-278-091-2
13	632	26.9	463	1	US-08-483-859-2
14	632	26.9	463	1	US-08-472-173-2
15	632	26.9	463	1	US-08-487-167-2
16	632	26.9	463	2	US-08-482-816-2
17	632	26.9	463	2	US-08-296-149-2
18	632	26.9	463	2	US-08-801-499-2
19	632	26.9	463	2	US-08-615-271-2
20	632	26.9	463	3	US-09-074-660-2
21	632	26.9	463	3	US-09-074-659-2
22	632	26.9	463	3	US-09-106-468-2
23	632	26.9	463	4	US-09-106-467-2
24	632	26.9	463	4	US-09-106-467-2
25	631.5	26.8	475	1	US-08-278-091-6
26	631.5	26.8	475	1	US-08-483-859-6
27	631.5	26.8	475	1	US-08-472-173-6

28	631.5	26.8	475	2	US-08-487-167-6	Sequence 6, Appl1
29	631.5	26.8	475	2	US-08-482-816-6	Sequence 6, Appl1
30	631.5	26.8	475	2	US-08-296-149-6	Sequence 6, Appl1
31	631.5	26.8	475	2	US-08-801-499-6	Sequence 6, Appl1
32	631.5	26.8	475	2	US-08-615-271-6	Sequence 6, Appl1
33	631.5	26.8	475	3	US-09-074-660-6	Sequence 6, Appl1
34	631.5	26.8	475	3	US-09-074-659-6	Sequence 6, Appl1
35	631.5	26.8	475	3	US-09-106-468-6	Sequence 6, Appl1
36	631.5	26.8	475	4	US-09-106-467-6	Sequence 6, Appl1
37	631.5	26.8	475	4	US-09-106-467-6	Sequence 6, Appl1
38	606.5	25.8	472	1	US-08-278-091-5	Sequence 5, Appl1
39	606.5	25.8	472	1	US-08-483-859-5	Sequence 5, Appl1
40	606.5	25.8	472	1	US-08-472-173-5	Sequence 5, Appl1
41	606.5	25.8	472	2	US-08-487-167-5	Sequence 5, Appl1
42	606.5	25.8	472	2	US-08-482-816-5	Sequence 5, Appl1
43	606.5	25.8	472	2	US-08-296-149-5	Sequence 5, Appl1
44	606.5	25.8	472	2	US-08-801-499-5	Sequence 5, Appl1
45	606.5	25.8	472	2	US-08-615-271-5	Sequence 5, Appl1

ALIGNMENTS

```
RESULT 1
US-09-199-637A-132
Sequence 132, Application US/09199637A
Patent No. 6355411
GENERAL INFORMATION:
APPLICANT: Ausubel, Frederick
APPLICANT: Goodman, Howard W.
APPLICANT: Rahme, Laurence G.
APPLICANT: Mahajan-Miklos, Shalina
APPLICANT: Tan, Man-Wah
APPLICANT: Cao, Hui
APPLICANT: Drenthard, Eliana
APPLICANT: Tsongalis, John
TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID
FILE REFERENCE: 00786/361002
CURRENT APPLICATION NUMBER: US/09/199, 637A
PRIORITY FILING DATE: 1998-11-25
PRIORITY APPLICATION NUMBER: 60/066, 517
PRIORITY FILING DATE: 1997-11-25
NUMBER OF SEQ ID NOS: 437
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 132
LENGTH: 460
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-199-637A-132
Query Match
Best Local Similarity 52.5%; Pred. No. 4,4e-84;
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;
54 LDPFAVLVSEGPVAVNI---QAAPARTONGSGNAETSDPLADSP-FIEFPKRLVFN 109
27 LDPFTLVQASAVVNISTRKLP---DRAARGLSTPDLGLPMPFDLERTIPQ 82
110 MPEIP--OEADGGLNFGSGFIISKNGYILNTNTHVAVAGSIKVLANDKREYAKLIG 166
83 VPRNPGRQGRQA---SLGSGFTIISNDGYILTNHVVADADEILVLSRSEHKAKLVG 138
167 SDVQSVALLKTDATLELPVVKIGNKNLKRGSVVAIGAPFGEDNSVTAGIVAKGRSL 226
139 ADPRSVAAVAKIEA-KNLPTLKGDNSNKLKGVSWVLAIGSPFGDHSVTAGIVAKGRSL 197
227 PMSYTPFTQTDVAINPNSGGLFNLKGVVGINSQIYRSQGFMSIPAIPIDVAMNV 286
198 PMSYTPFTQTDVAINPNSGGLNLBGEVVGINSQIFRSQGFMSIPAIPIDVAMNV 257
287 ARLQKTVGVQRCGLVITIQEVSYGIAQSGFLDKASGALITAKLPGSPARAGIQAGDIV 34
```

Db 258 ADOLKAGKVSRLGVLVIOEVNKLAESEGLDKPSGALVAQLVEDGPAKGLQVGDVI 317
QY 347 LSLDGEIRSGDLPVWVGATITPGKEVSLGVMRKGEEITIKAKLNAEHTGASSKTDE- 405
Db 318 LSLNOSINSEADLPVHVGMRKGDKNLIVIRNGQRKSLSMVGNLBD-----DDEE 370
QY 406 ----APYTEOQSGTFSVESAGITLQTH--TDSGKHLVVVRVSDA--AERAGLRHGEI 456
Db 371 IASMGAPGERSNRIGVTVAADLTAEQRKSLDIQG--GVVKEVQDGPAAVIGLRGDIVI 428

RESULT 2

US-08-245-294-8
Sequence 8, Application US/08245294
Patent No. 5644047

GENERAL INFORMATION:

APPLICANT: Anderson, Burt E.
APPLICANT: Regnery, Russell L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: DIAGNOSING
TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA
TITLE OF INVENTION: INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/245,294

FILING DATE:

CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:

NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,612
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 503 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-245-294-8

Query Match 31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. No. 7.8e-60;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCALIACGCEKASFFGADKKEASFYERIEHTKDDGSVSMILPDPFQVLOSEGPA 67
Db 15 FSAALFETALFFSC-----GSSLMTTKAHANSV-----FSSLMOQGGFADIVSQVKPA 62
QY 68 VNIQAAPARTON---GSGNAETDSDPLADSDP---FYEFFKRLVPMMEIIPQEBADD 120
Db 63 VSVQVKKKKKKKKKEMPPSPDFSTPGFQDLPDQHLKXFPQDFYNRDKFSNKSIL-QRSHRL 121
QY 121 GGLNFGSGFIISKNGYILTNTTHVAVGNGSIKVLLNDRREYTKLIGSDVQSDVALLKIDA 180
Db 122 RIARSGSGFISDGYIVTNNHVISDGSYAIVVLDGTELNAKLIGTDPRTDLAVLKVNE 181
QY 181 TBELPVYKGNPNKLRGEEVVAALIGAPFGPDNSTAGIVSAKGSLLPMESTTFPIQTIVA 240
Db 182 KRKFSYVDEGDSKLRVGDVVAALIGNPFGLGTAGIVSARGRDICTGVYDDFIQIDAA 241

QY 241 INPNSGGLPFLNKGQVVGINSQIYSRSRSGFMGISFAIPIDVANNVAEQLNKTGKVRQO 300
Db 242 VNRNGSGFTFDLNGKVVGVNTAIFSPSGNVGIAFAIPATVAHEVQQLIEKGLVORGM 301
QY 301 LGVITIOEVSTGLAQSGLDPAKSGALLIAKIIPGSAERAGLOAGDIVLSLDGEIRSSGL 360
Db 302 LGVQIOPVTKESIDSLGKKAQALITDPLKG--PAARAGIRAGDIVISVNGEKINDVRDL 360
QY 361 PWWVGATITPGKEVSLGVMRKGEEITIKAKLNAEHTGASSKTDEAPYT--EQQSGTFSVE 419
Db 361 AKRIANMSRGETYTLTGVMKSGKEENIKYKLDSPED---ENMKDQSKTSNEGNNDITLE 417
QY 420 SAGITLQTHDSSGKHLVVVRV---SDAERAGLRHGEIILAV 459
Db 418 DGLIYAPSDDGLG--LVVTVDVDPDSDAADK-GIRPGDIVITV 457

RESULT 3

US-08-474-499-8
Sequence 8, Application US/08474499
Patent No. 5693776

GENERAL INFORMATION:

APPLICANT: Anderson, Burt E.
APPLICANT: Regnery, Russell L.
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR
TITLE OF INVENTION: DIAGNOSING ROCHALIMAEA HENSELAE
TITLE OF INVENTION: AND ROCHALIMAEA QUINTANA INFECTION
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSER: NEEDLE & ROSENBERG, P.C.
STREET: 127 Peachtree Street, Suite 1200
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/474,499
FILING DATE: 07-JUN-1995
CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/245,294
FILING DATE: 18-MAY-1994

ATTORNEY/AGENT INFORMATION:

NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,612
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein

US-08-474-499-8

Query Match 31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. No. 7.8e-60;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCALIACGCEKASFFGADKKEASFYERIEHTKDDGSVSMILPDPFQVLOSEGPA 67
Db 15 FSAALFETALFFSC-----GSSLMTTKAHANSV-----FSSLMOQGGFADIVSQVKPA 62
QY 68 VNIQAAPARTON---GSGNAETDSDPLADSDP---FYEFFKRLVPMMEIIPQEBADD 120

```

Db      63 VVSQVQSKNKKKKWFSDFFSTPGFDQLPDQHPLKFFQDFYNRDKPSNKSU-ORSHRL 121
Qy      121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIGSDVQSPVALLKIDA 180
      122 RPIAFSGGFIISSDGIYVTNNHVISDGTSYAVVLDGTELMNAKLIGTDPRTDLAVLKVNE 181
Qy      181 TEELPVVKIKGNPKNLKPGEWVAAGAPFGFNDSTAGIVSAKGRSLPNESYTPFIQTDA 240
      182 KRKFSYVDFGDDSKLRVGDWVAAGNPFGLGTVTAGIVSARGRDIGTVYDDFIQIDA 241
Qy      241 INPNSGSGPLFNLKGQVVGINSQIYSRSGFMGISFAIPDIVAMVAEQLKATGKVRQ 300
      242 VNRGNSGPTFDNLKGVGVNTAIFSPSGNVGIAFAIPATANEVQQLIEKGLVQRGW 301
Qy      301 LGVIOEVSYGIAQSGFGLDKASGALIAKILPGSPAERAGLOAGDIVLSLDDGIRSSGDL 360
      302 LGVIOPTVTKIISDSIGLKEAKGALITDPLKG-PAKAKGIRAGDVIIISVNGEKINDVRL 360
Qy      361 PVMVGAITPKEKESLGVWRKGEIITIKAKLGNAAEHTGASKTDEAPYT-EOQSGTFSVE 419
      361 AKRIANMSPGETVTLGVWMSGKEENIKVLDSPED--ENMKDGSKYSNEHNSDETLE 417
Qy      420 SAGITLQTHTSSGKHLVVVRV---SDAERAGLRHGDILLAV 459
      418 DYGLIVAPSDDLG--LVVTVDVDPDSDAADK-GIRPGDIVITV 457
Db

```

RESULT 4

```

US-08-307-279A-8
; Sequence 8, Application US/08307279A
; Patent No. 5736347
; GENERAL INFORMATION:
; APPLICANT: Anderson, Burt E.
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
; TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/307,279A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-307-279A-8

```

```

Query Match      31.1%; Score 731.5; DB 1; Length 503;
Best Local Similarity 39.5%; Pred. No. 7,8e-50;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

```

```

Qy      8 FALAALCAALLAGECEKAGSFFGADKKEASFVERIEHTKDDGSVSNMLDPFAQLVQSEGA 67
      15 FSALETFALFFSGC---GSSIMTTKHAANSY-----FSSIMQOQGFADIVSQVKA 62
Qy      68 VVNIQAAPARTQN---GSGNAETDSPPLADSP----FYFFPKRLVNMPEIPQBEADD 120
      63 VVSQVQSKNKKKKWFSDFFSTPGFDQLPDQHPLKFFQDFYNRDKPSNKSU-ORSHRL 121
Qy      121 GGLNFGSGFIISKNGYILTNTHVAVAGMSIKVLLNDKREYAKLIGSDVQSPVALLKIDA 180
      122 RPIAFSGGFIISSDGIYVTNNHVISDGTSYAVVLDGTELMNAKLIGTDPRTDLAVLKVNE 181
Qy      181 TEELPVVKIKGNPKNLKPGEWVAAGAPFGFNDSTAGIVSAKGRSLPNESYTPFIQTDA 240
      182 KRKFSYVDFGDDSKLRVGDWVAAGNPFGLGTVTAGIVSARGRDIGTVYDDFIQIDA 241
Qy      241 INPNSGSGPLFNLKGQVVGINSQIYSRSGFMGISFAIPDIVAMVAEQLKATGKVRQ 300
      242 VNRGNSGPTFDNLKGVGVNTAIFSPSGNVGIAFAIPATANEVQQLIEKGLVQRGW 301
Qy      301 LGVIOEVSYGIAQSGFGLDKASGALIAKILPGSPAERAGLOAGDIVLSLDDGIRSSGDL 360
      302 LGVIOPTVTKIISDSIGLKEAKGALITDPLKG-PAKAKGIRAGDVIIISVNGEKINDVRL 360
Qy      361 PVMVGAITPKEKESLGVWRKGEIITIKAKLGNAAEHTGASKTDEAPYT-EOQSGTFSVE 419
      361 AKRIANMSPGETVTLGVWMSGKEENIKVLDSPED--ENMKDGSKYSNEHNSDETLE 417
Qy      420 SAGITLQTHTSSGKHLVVVRV---SDAERAGLRHGDILLAV 459
      418 DYGLIVAPSDDLG--LVVTVDVDPDSDAADK-GIRPGDIVITV 457
Db

```

RESULT 5

```

US-09-525-310-8
; Sequence 8, Application US/09525310
; Patent No. 6406887
; GENERAL INFORMATION:
; APPLICANT: Regnery, Russell L.
; TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae
; TITLE OF INVENTION: and Methods and Compositions for Diagnosing Rochalimaea
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.
; STREET: 127 Peachtree Street, N.E., Suite 1200
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/525,310
; FILING DATE: 14-Mar-2000
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/307,279
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Spratt, Gwendolyn D.
; REGISTRATION NUMBER: 36,016
; REFERENCE/DOCKET NUMBER: 1414.624
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 688-0770
; TELEFAX: (404) 688-9880
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 503 amino acids

```

TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 SEQUENCE DESCRIPTION: SEQ ID NO: 8:
 US-09-525-310-8

Query Match 31.1%; Score 731.5; DB 4; Length 503;
 Best Local Similarity 39.5%; Pred. No. 7.8e-60;
 Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCAALLACCEKAGSFFGADKKEASFVERIEHTKDQGSVSMLEPFAQLVQSEGA 67
 DB 15 FSALETLFFSGC---GSSLMTTKAHANSV-----FSSLMOQOGFADIVSQVKPA 62
 QY 68 VVNIQAPAPRTQN---GSGNAETDSPLADSDP----FYEFKRLVPMNPEIPOEADD 120
 DB 63 VVSQVQKSNKKKKEMFFSDPFSTPGFDLPDQHPLKFFQDFYNRDKPSNKSLL-QRSRL 121
 QY 121 GGLNFGSGFTISKNGYLLTNTHVVAAGMSIKVLLNDKREYTKLIGSDVQSDVALLKIDA 180
 DB 122 RPIAFSGSFFIISDGYIVTNHVISDGTSYAVVLDGTELNAKLIGTDPRTDLAVLKVNE 181
 QY 181 TELPLPVYKGNPKLKGEMVVAAGAPFGPDNSTAGIVSAKGRSLPNESTPFIQTDVA 240
 DB 182 KRKFSYVDFGDDSKLRVGDWVAALGNPFGAGITVAGIVSARGDITGVYDDFIQIDAA 241
 QY 241 INPNSGGPLFNLKQGVGINSQIYSRSGFGMISFAIPIDVAMNVAEQLKNTGKVQRGQ 300
 DB 242 VNRGNSGGPTFDLNGKVVGVNTAIFSPSGGNVGIAPAIIPAATANEVVOQLIEKGLVQRGW 301
 QY 301 LGVITIOEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLOAGDIVLSLGGEIRSSGDL 360
 DB 302 LGVOIQPTKEISDSIGLKEAKGALITDPLKG--PAKAKIGIKAGDIVIISVNGEKINDVRL 360
 QY 361 PVMVGAITPGKEVSLGVRKGEETITKAKIGNAEHTGASSTDEAPYT--EQSGTSFVSE 419
 DB 361 AKRIANMSPGETVTLGVWKSCKEENIKVLDSPED--ENMKDQSKYSNEHNSDETLE 417
 QY 420 SAGITLQTHTDSGKGLVWVRV---SDAERAGLRHGEIILAV 459
 DB 418 DYGLIVAPSDDGLG--LVVTVDVDPDSDADK-GIRPGDIVTV 457

RESULT 6
 PCT-US95-06211-8
 ; Sequence 8, Application PC/TUS9506211
 ; GENERAL INFORMATION:

APPLICANT:
 TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
 TITLE OF INVENTION: ROCHALIMAEA HENSELAE AND ROCHALIMAEA QUINTANA INFECTION
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: NEEDLE & ROSENBERG, P.C.
 STREET: 127 Peachtree Street, Suite 1200
 CITY: Atlanta
 STATE: Georgia
 COUNTRY: USA
 ZIP: 30303
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: PCT/US95/06211
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/245,294
 FILING DATE: 18 MAY 1994
 CLASSIFICATION:
 ATTORNEY/AGENT INFORMATION:
 NAME: Spratt, Gwendolyn D.
 REGISTRATION NUMBER: 36,016
 REFERENCE/DOCKET NUMBER: 1414.6121

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 404/688-0770
 TELEFAX: 404/688-9880
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 503 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 PCT-US95-06211-8

Query Match 31.1%; Score 731.5; DB 5; Length 503;
 Best Local Similarity 39.5%; Pred. No. 7.8e-60;
 Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

QY 8 FALAALCAALLACCEKAGSFFGADKKEASFVERIEHTKDQGSVSMLEPFAQLVQSEGA 67
 DB 15 FSALETLFFSGC---GSSLMTTKAHANSV-----FSSLMOQOGFADIVSQVKPA 62
 QY 68 VVNIQAPAPRTQN---GSGNAETDSPLADSDP----FYEFKRLVPMNPEIPOEADD 120
 DB 63 VVSQVQKSNKKKKEMFFSDPFSTPGFDLPDQHPLKFFQDFYNRDKPSNKSLL-QRSRL 121
 QY 121 GGLNFGSGFTISKNGYLLTNTHVVAAGMSIKVLLNDKREYTKLIGSDVQSDVALLKIDA 180
 DB 122 RPIAFSGSFFIISDGYIVTNHVISDGTSYAVVLDGTELNAKLIGTDPRTDLAVLKVNE 181
 QY 181 TELPLPVYKGNPKLKGEMVVAAGAPFGPDNSTAGIVSAKGRSLPNESTPFIQTDVA 240
 DB 182 KRKFSYVDFGDDSKLRVGDWVAALGNPFGAGITVAGIVSARGDITGVYDDFIQIDAA 241
 QY 241 INPNSGGPLFNLKQGVGINSQIYSRSGFGMISFAIPIDVAMNVAEQLKNTGKVQRGQ 300
 DB 242 VNRGNSGGPTFDLNGKVVGVNTAIFSPSGGNVGIAPAIIPAATANEVVOQLIEKGLVQRGW 301
 QY 301 LGVITIOEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLOAGDIVLSLGGEIRSSGDL 360
 DB 302 LGVOIQPTKEISDSIGLKEAKGALITDPLKG--PAKAKIGIKAGDIVIISVNGEKINDVRL 360
 QY 361 PVMVGAITPGKEVSLGVRKGEETITKAKIGNAEHTGASSTDEAPYT--EQSGTSFVSE 419
 DB 361 AKRIANMSPGETVTLGVWKSCKEENIKVLDSPED--ENMKDQSKYSNEHNSDETLE 417
 QY 420 SAGITLQTHTDSGKGLVWVRV---SDAERAGLRHGEIILAV 459
 DB 418 DYGLIVAPSDDGLG--LVVTVDVDPDSDADK-GIRPGDIVTV 457

RESULT 7
 US-08-485-569-2
 ; Sequence 2, Application US/08485569
 ; Patent No. 5679547
 ; GENERAL INFORMATION:
 APPLICANT: Krivan, Howard C.
 APPLICANT: Samuel, James E.
 TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
 TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
 NUMBER OF SEQUENCES: 2
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of Americas
 CITY: New York
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 10036
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/485,569
 FILING DATE: 07-JUN-1995

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/903,079
FILING DATE: 22-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,966
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/631,698
FILING DATE: 21-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Baldwin, Geraldine F.
REGISTRATION NUMBER: 31,232
REFERENCE/DOCKET NUMBER: 7969-049
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-485-569-2

Query Match 27.5%; Score 646; DB 1; Length 463;
Best Local Similarity 38.9%; Pred. No. 6,6e-52;
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;

QY 54 LPDPAQLVQSEGPAVNIQAAPARTONGSNAETDS--DPLASDP--FYEFF-KRLVFN 109
DB LPSPFVSQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSRSPFLDIPBEKFFFGDRFAEQ 85
QY 110 MEIIPQEAADGGNLFSGFII-SKNGYILTNTHVVAAGMSIKVLLNDRKREYAKLIGSD 168
DB 86 FGGRGSKRNFRGL--GSGVTIINASKGYVLTNNHVIDGADKITVQLQDGRBFKAKLVGKD 143
QY 169 VQSDVALLKIDATEELPVVKIGNPKNLKPEGVAAIGAPFGFDSVTAGIVSAKGRSLPN 228
DB 144 EQSDIALVQLEKPSNLTEIKFADSDKLVGDFTVAINGNPGLGQVTSIGIVSALGRSTGS 203
QY 229 ES--YTFPIQTVAIINPGNSGGLFNLKGQVAINSGIYRSRSGFMGISFALPIDVMNV 286
DB 204 DSGTYENYIQTDAVNRGNSGGLVNLNGELIGINTALISPSGNAAGIAFAIPNSQASNL 263
QY 287 AEOLKNTKXVORGLGVIIQEVSYGLAOSFGLDKASGALIAKILFSGPAERGLQAGDIY 346
DB 264 VQQLLEFGVRRGLGIGKSELNADLAKAFNVSAQOGAFVSEVLPKSAAEKAKGLAKAGDII 323
QY 347 LSLDGEIRSSGDLPMVGAITTPGKEVSLGWRKGEITTIKALGNAAEHTGASSKTDEA 406
DB 324 TANNGOKISSFAEIRAKIATTGAKKEISLTVLRDGRSHVPMKL-QADDSGLSSKT-EL 381
QY 407 PYTEQSGTFSVESAGITLTQHTDSSGKHLVVVRV--SDAERAGLRHGEDEIIV 459
DB 382 PALD-----GATLKDYDAKGVKGIETIKIOPNSLAQR-GLKSGDIIIGI 425

RESULT 8
US-08-480-993-2
Sequence 2, Application US/08480993
Patent No. 572115
GENERAL INFORMATION:
APPLICANT: Kriyvan, Howard C.
APPLICANT: Samuel, James E.
APPLICANT: No. 572115BEG, Nils T.
TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pemle & Edmonds

STREET: 1155 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/480,993
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/903,079
FILING DATE: 22-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,966
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/631,698
FILING DATE: 21-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7969-050
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-480-993-2

Query Match 27.5%; Score 646; DB 1; Length 463;
Best Local Similarity 38.9%; Pred. No. 6,6e-52;
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;

QY 54 LPDPAQLVQSEGPAVNIQAAPARTONGSNAETDS--DPLASDP--FYEFF-KRLVFN 109
DB LPSPFVSQNSLAPMLEKVQ--PAVVTLSVEGKAKVDSRSPFLDIPBEKFFFGDRFAEQ 85
QY 110 MEIIPQEAADGGNLFSGFII-SKNGYILTNTHVVAAGMSIKVLLNDRKREYAKLIGSD 168
DB 86 FGGRGSKRNFRGL--GSGVTIINASKGYVLTNNHVIDGADKITVQLQDGRBFKAKLVGKD 143
QY 169 VQSDVALLKIDATEELPVVKIGNPKNLKPEGVAAIGAPFGFDSVTAGIVSAKGRSLPN 228
DB 144 EQSDIALVQLEKPSNLTEIKFADSDKLVGDFTVAINGNPGLGQVTSIGIVSALGRSTGS 203
QY 229 ES--YTFPIQTVAIINPGNSGGLFNLKGQVAINSGIYRSRSGFMGISFALPIDVMNV 286
DB 204 DSGTYENYIQTDAVNRGNSGGLVNLNGELIGINTALISPSGNAAGIAFAIPNSQASNL 263
QY 287 AEOLKNTKXVORGLGVIIQEVSYGLAOSFGLDKASGALIAKILFSGPAERGLQAGDIY 346
DB 264 VQQLLEFGVRRGLGIGKSELNADLAKAFNVSAQOGAFVSEVLPKSAAEKAKGLAKAGDII 323
QY 347 LSLDGEIRSSGDLPMVGAITTPGKEVSLGWRKGEITTIKALGNAAEHTGASSKTDEA 406
DB 324 TANNGOKISSFAEIRAKIATTGAKKEISLTVLRDGRSHVPMKL-QADDSGLSSKT-EL 381
QY 407 PYTEQSGTFSVESAGITLTQHTDSSGKHLVVVRV--SDAERAGLRHGEDEIIV 459
DB 382 PALD-----GATLKDYDAKGVKGIETIKIOPNSLAQR-GLKSGDIIIGI 425

RESULT 9

US-07-903-079B-2
Sequence 2, Application US/07903079B
Patent No. 5843463
GENERAL INFORMATION:
APPLICANT: Kriivan, Howard C.
APPLICANT: Samuel, James E.
APPLICANT: No. 5843463berg, Nile T.
TITLE OF INVENTION: ADHESIN-OLIGOSACCHARIDE CONJUGATE
TITLE OF INVENTION: VACCINE FOR HAEMOPHILUS INFLUENZAE
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/903,079B
FILING DATE: 22-JUN-1992
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/810,966
FILING DATE: 20-DEC-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/631,698
FILING DATE: 21-DEC-1990
ATTORNEY/AGENT INFORMATION:
NAME: Mastrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 7969-004
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8664/9741
TELEX: 66141 PENNTE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
7-903-079B-2

Query Match 27.5%; Score 646; DB 2; Length 463;
Best Local Similarity 38.9%; Pred. No. 6,6e-52;
Matches 162; Conservative 75; Mismatches 151; Indels 28; Gaps 12;
54 LPDPAQLVQSEGPVVNIQAAPRTQSGNAGTDS-DPLADSDP--FYEFP-KRLVNV 109
28 LPSFVSEQNSLAPMLRKVQ--PAVVTLSVEGKAKVDSRSPFLDDIPEEFKFFGDRFAEQ 85
110 MEIITPOEADDDGLNFGSGFIT-SKNGYILTNTHTYVAVGMSIKVLNDKREYTAQLISD 168
86 FGGRESKSNFRGL--GSGVITINASKGYVLTNNHVIDGAKITVQLDGRBFKALVQKD 143
169 VQSDVALLKIDATEELPVVKGPNKMLKPGEMWVAIGAPFGFDSVTAGIYSAKGRSLPN 228
144 EOSDIALVOLERKSNLTETKFAVSDKLRVDFVYALGNPFLGQTVTSIGIVSALGRSTGS 203
229 ES--VTFPIQTVDVAINPGNSGGLPFLKGGVGVINSQIYSRSGSGFMGISFAIPIDVANNV 286
204 DSGTYENYIQTDAAVVRNGSGALVNLNGELIGINTALISPSGNGAGIAFAIPNSQASNL 263
287 AEQLKNTGVQNGQGLVITIQEVSYGLAQSGFLDKASGALLAKITPGSPAERAGLQAGIV 346
264 VQVILFEQVRRGLLGIKGGELNADLAKAFNVSAQOGAFVSEVLPKSAAERAGLKAQDI 323

QY 347 LSLDGEIRSSGDLPVWVGAITPKEVSLGWRKGEIITIKAGNAAEHTGASSKTDEA 406
DB 324 TAWNGKISSFARIRAKIATGAGKEISLFTYLRGKSHDVWKL-QADDSGLSSXT-EL 381
QY 407 PYTEQSGTFSVESAGITTLQTHDSSGKLVVAV---SDAARAGLRHGEITLAV 459
DB 382 PALD-----GATLNDYDAKGVGKIEITRIQPNSLAQR-GLKSGDITIIIGI 425

RESULT 10

US-08-350-741-2
Sequence 2, Application US/08350741
Patent No. 5804194
GENERAL INFORMATION:
APPLICANT: DOUGAN G.
APPLICANT: CHARLES I.G.
APPLICANT: HORMACHE C.E.,
APPLICANT: JOHNSON K.S.,
APPLICANT: CHATFIELD S.N.
TITLE OF INVENTION: LIVE VACCINES
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: NIXON and VANDERHYE PC
STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22201-4714
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/350,741
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/952,737
FILING DATE: 09-MAY-1994
APPLICATION NUMBER: US 07/952,737
FILING DATE: 30-NOV-1992
APPLICATION NUMBER: GB 9007194.5
FILING DATE: 30-MAR-1990
APPLICATION NUMBER: PCT/GB91/00484
FILING DATE: 28-MAR-1991
ATTORNEY/AGENT INFORMATION:
NAME: WILSON, MARY J.
REGISTRATION NUMBER: 32,955
REFERENCE/DOCKET NUMBER: 117-158
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 816-4000
TELEFAX: (703) 816-4100
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 475 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-350-741-2

Query Match 26.9%; Score 633.5; DB 1; Length 475;
Best Local Similarity 34.1%; Pred. No. 1e-50;
Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;
QY 44 TXDDGSVSMLLPFAQLVQSEGPVV--NIQADAPRTQSGNAGTDS--TDSPLADSD 97
DB 29 TSSANTAOQMPLAPLMEKVMPSVVSINVGSTTVTPRPRRPNQGFQDSDSPFCQDGS 88
QY 98 PYEFPKRLVPMNPEIPOEADQG-----LNFSGFTI-SKNGYILTNTHTYVAVGK 148

Db 89 PF-----QNSPFC--QGCGNGNGGQOQKFMALGSGVITIDAAQYVVTNNHVVDNAS 138
 QY 149 SIKVLNDRKREYTAKLIGSDVQSDVALLKIDATEELPVYKIGNPKNLKPGEWAAIGAPF 208
 Db 139 VIKVQLSDGRKFDKAVGKDPKRSIDIALIQIONPKNLTAIKLADSDALRVGDYVAIGNPF 198
 QY 209 GFDNSVTAGIVSAKGRS-LPNESYTPFIOTDVAINPGNSGGPLFNLKGOVVGINSQIYSR 267
 Db 199 GLGETVTSIGIVSALGRSGLVNENYENFIOTDVAINRNGSGALVNLNGELIGINTAILAP 258
 QY 268 SGGFMGISFAIPIDVAMNVAEQLKNTGKVGORGLGVIIEVSYGLAOSFGLDKASGALIA 327
 Db 259 DCGNIGIGFALPISNMVKNLTSQWVEYGVKRGELIGMTELSLAKAMKVDKQAGAFVS 318
 QY 328 KILGSPAPERAQLOAGDIVLSLDGGEIRSSGDLPMVVGAIITPGKEVSLGVWRKGEIITIK 387
 Db 319 QVMPNSSAAKAGIKAGDVITSLNGKPISSFPALRAQVGTMPVGSKISLGLRKGALITVN 378
 QY 388 AKLGNAEHHTGASSKTDEAPYTEQSGTSFVESA-----GITLQTHTDSGKHLVYVR 440
 Db 379 LEL-----QQSSQSQVDSSTIFSGIEGAEKSNKGQDKGVVSVYK 418
 QY 441 VSDAERAGLRHGDIELAVRASP 463
 Db 419 ANSPAAQIGLKKGDVITIGANQOP 441

RESULT 11
 US-08-463-875A-2
 ; Sequence 2, Application US/08463875A
 ; Patent No. 5980907
 ; GENERAL INFORMATION:
 ; APPLICANT: DOUGAN, Gordon
 ; APPLICANT: CHARLES, Ian G.
 ; APPLICANT: HORMACHE, Carlos E.
 ; APPLICANT: JOHNSON, Kevin S.
 ; APPLICANT: CHATFIELD, Steven N.
 ; TITLE OF INVENTION: LIVE VACCINES
 ; NUMBER OF SEQUENCES: 3
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: NIXON and VANDERHYE PC
 ; STREET: 8th FLOOR, 1100 NORTH GLEBE ROAD
 ; CITY: ARLINGTON
 ; STATE: VIRGINIA
 ; COUNTRY: USA
 ; ZIP: 22201-4714
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/463.875A
 ; FILING DATE: 05-JUN-1995
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US 07/340,741
 ; FILING DATE: 07-DEC-1994
 ; APPLICATION NUMBER: US 07/952,737
 ; FILING DATE: 30-NOV-1992
 ; APPLICATION NUMBER: GB 9007194.5
 ; FILING DATE: 30-MAR-1990
 ; APPLICATION NUMBER: PCT/GB91/00484
 ; FILING DATE: 28-MAR-1991
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: WILSON, MARY J.
 ; REGISTRATION NUMBER: 32,955
 ; REFERENCE/DOCKET NUMBER: 117-158
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 816-4100
 ; TELEFAX: (703) 816-4100
 ; TELEX: 200797 NIXN UR
 ; INFORMATION FOR SEQ ID NO: 2:
 ; SEQUENCE CHARACTERISTICS:

; LENGTH: 475 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-463-875A-2

Query Match 26.9%; Score 633.5; DB 2; Length 475;
 Best Local Similarity 34.1%; Pred. No. 1e-50;
 Matches 151; Conservative 83; Mismatches 156; Indels 53; Gaps 9;

QY 44 TXDDGSVSMLLPDPFAQVQSBGPAVV--NIOAAAPRTQNSGNNE-----TDSPLASD 97
 Db 29 TSSSAMTAQOMFSLAPLETKVMPBSVINVEGSTTVTPRMPRFQGFPGDSFPCOQGS 88
 QY 98 PFYEFFKRLVPMNPEIQEADQD-----LNFSGFII-SKNGYITLNTHTVAVAGM 148
 Db 89 PF-----QNSPFC--QGCGNGNGGQOQKFMALGSGVITIDAAKAGIVVTNNHVVDNAS 138
 QY 149 SIKVLNDRKREYTAKLIGSDVQSDVALLKIDATEELPVYKIGNPKNLKPGEWAAIGAPF 208
 Db 139 VIKVQLSDGRKFDKAVGKDPKRSIDIALIQIONPKNLTAIKLADSDALRVGDYVAIGNPF 198
 QY 209 GFDNSVTAGIVSAKGRS-LPNESYTPFIOTDVAINPGNSGGPLFNLKGOVVGINSQIYSR 267
 Db 199 GLGETVTSIGIVSALGRSGLVNENYENFIOTDVAINRNGSGALVNLNGELIGINTAILAP 258
 QY 268 SGGFMGISFAIPIDVAMNVAEQLKNTGKVGORGLGVIIEVSYGLAOSFGLDKASGALIA 327
 Db 259 DCGNIGIGFALPISNMVKNLTSQWVEYGVKRGELIGMTELSLAKAMKVDKQAGAFVS 318
 QY 328 KILGSPAPERAQLOAGDIVLSLDGGEIRSSGDLPMVVGAIITPGKEVSLGVWRKGEIITIK 387
 Db 319 QVMPNSSAAKAGIKAGDVITSLNGKPISSFPALRAQVGTMPVGSKISLGLRKGALITVN 378
 QY 388 AKLGNAEHHTGASSKTDEAPYTEQSGTSFVESA-----GITLQTHTDSGKHLVYVR 440
 Db 379 LEL-----QQSSQSQVDSSTIFSGIEGAEKSNKGQDKGVVSVYK 418
 QY 441 VSDAERAGLRHGDIELAVRASP 463
 Db 419 ANSPAAQIGLKKGDVITIGANQOP 441

RESULT 12
 US-08-278-091-2
 ; Sequence 2, Application US/08278091
 ; Patent No. 5506139
 ; GENERAL INFORMATION:
 ; APPLICANT: LOOSMORE, Sheena M
 ; APPLICANT: YANG, Yan-Ping
 ; APPLICANT: CHONG, Pele
 ; APPLICANT: OOMEN, Raymond P.
 ; APPLICANT: KLEIN, Michel H.
 ; TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
 ; NUMBER OF SEQUENCES: 23
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Sim & McBurney
 ; STREET: Suite 701, 330 University Avenue
 ; CITY: Toronto
 ; STATE: Ontario
 ; COUNTRY: Canada
 ; ZIP: M5G 1R7
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/278,091
 ; FILING DATE: 21-JUL-1994
 ; CLASSIFICATION: 435
 ; ATTORNEY/AGENT INFORMATION:

NAME: Stewart, Michael I
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-371
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-278-091-2

Query Match 26.9%; Score 632; DB 1; Length 463;
Best Local Similarity 38.5%; Pred. No. 1.3e-50;
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

QY 54 LPDFALVQSEGPAAVNIQAAPAPRTQNGSGNAETDS--DPLADSDP--FYEFF-KRLVFN 109
28 LPSFVSEONSLAPMLEKVQ--PAVVTLSVEGAKVDSDRSPLDDIPEEFKFFGDRFAEQ 85
110 MPEIPEQEBADDDGILNFGSGFII-SKNGYILTNTHVAVMGSIKVLNDKREYAKLIGSD 168
DB 86 FGGRGESKRNFRGL--GSGVITINASKGYVLTNNHVIDEADKITVQLQDGRREFRAKLVGKD 143
QY 169 VQSDVALKIDATEELPVVVKIGNPKNLKPGEVVAIAGAPFGDNTYAGIVSAKGRSLRN 228
DB 144 ELSDIALVQLEKSPNLTETIKFADSDKLRVGDFTVAIGNPGLQGTVTSIGIVSALGRSTGS 203
QY 229 ES--YTFPIOTDVAINPNSGGPLFNLKGOVVGINSQIYRSRGGFMGISFAIPIDVAMNV 286
DB 204 DSGTYENYIOTDVAANVRNNGSGALVNLNGELIGINTAIISPSGNAIGAFAPISNOASNL 263
QY 287 AEOLKNTKVGQVQGLVILIQEVSYGLAOSFGLDKASGALIAKILPGSPARAGLOAGDIY 346
DB 264 VQQLIEFGQVRKRLGLIGKSELNADLAKAFNVSAQOGAFVSEVLPKSAAEKXGLKAGDII 323
QY 347 LSLDGEIRSSGDLPVWVGAIITPGEVSLGVRKGEBITIKAKLGNAAEHTGASSTDEA 406
DB 324 TAMNGQKISSFAEIRAKIATTGAGKEISLTYLRDCKSHDVKKKL-QADDSQQLSSKT-EL 381
QY 407 PYTEQSGTFSVSAGITLQTHTDSSGKHLVVVR--SDAERAGLRHGEITLAV 459
DB 382 PALD-----GATLKDYDAKGVKGIETIKIOPNSLAQR-GLKSGDIIIGI 425

RESULT 13
US-08-483-859-2
Sequence 2, Application US/08483859
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: COMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/483,859
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET NUMBER: 1038-495 MIS.VG
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-483-859-2

Query Match 26.9%; Score 632; DB 1; Length 463;
Best Local Similarity 38.5%; Pred. No. 1.3e-50;
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

QY 54 LPDFALVQSEGPAAVNIQAAPAPRTQNGSGNAETDS--DPLADSDP--FYEFF-KRLVFN 109
DB 28 LPSFVSEONSLAPMLEKVQ--PAVVTLSVEGAKVDSDRSPLDDIPEEFKFFGDRFAEQ 85
QY 110 MPEIPEQEBADDDGILNFGSGFII-SKNGYILTNTHVAVMGSIKVLNDKREYAKLIGSD 168
DB 86 FGGRGESKRNFRGL--GSGVITINASKGYVLTNNHVIDEADKITVQLQDGRREFRAKLVGKD 143
QY 169 VQSDVALKIDATEELPVVVKIGNPKNLKPGEVVAIAGAPFGDNTYAGIVSAKGRSLRN 228
DB 144 ELSDIALVQLEKSPNLTETIKFADSDKLRVGDFTVAIGNPFGIQTTSIGIVSALGRSTGS 203
QY 229 ES--YTFPIOTDVAINPNSGGPLFNLKGOVVGINSQIYRSRGGFMGISFAIPIDVAMNV 286
DB 204 DSGTYENYIOTDVAANVRNNGSGALVNLNGELIGINTAIISPSGNAIGAFAPISNOASNL 263
QY 287 AEOLKNTKVGQVQGLVILIQEVSYGLAOSFGLDKASGALIAKILPGSPARAGLOAGDIY 346
DB 264 VQQLIEFGQVRKRLGLIGKSELNADLAKAFNVSAQOGAFVSEVLPKSAAEKXGLKAGDII 323
QY 347 LSLDGEIRSSGDLPVWVGAIITPGEVSLGVRKGEBITIKAKLGNAAEHTGASSTDEA 406
DB 324 TAMNGQKISSFAEIRAKIATTGAGKEISLTYLRDCKSHDVKKKL-QADDSQQLSSKT-EL 381
QY 407 PYTEQSGTFSVSAGITLQTHTDSSGKHLVVVR--SDAERAGLRHGEITLAV 459
DB 382 PALD-----GATLKDYDAKGVKGIETIKIOPNSLAQR-GLKSGDIIIGI 425

RESULT 14
US-08-472-173-2
Sequence 2, Application US/08472173
Patent No. 565353
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.
APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: COMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hin47 Protein with
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & McBurney
STREET: Suite 701, 330 University Avenue

CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,173
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET INFORMATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-472-173-2

Query Match 26.9%; Score 632; DB 1; Length 463;
Best Local Similarity 38.5%; Pred. No. 1.3e-50;
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

QY 54 LPDPAQLVQSEGPVAVNIQAAPARTONGSGNAETDS-DPLADSDP--FYEPF-KRLVYN 109
DB 28 LPSPVSEQNSLAPMLEKVQ--PAVVTLSVSGKAKVDSRSPFLDDIPEEFKFFFGDRFAEQ 85
QY 110 MPEIPOEADDDGLNFGSGFII-SKNGYILTNTHVAVGMGSIKYLLNDKREYTAKLIGSD 168
DB 86 FGGRGESKRNFRGL--GSGVILNASKGYVLTNNHVVIDEADKITVQLODGRERAKLVGKD 143
QY 169 VQSDVALLKIDATEELPVVKIKNPKNLKPGEWAAIGAPFFGDNSTYAGIVSAKRSPLN 228
DB 144 ELSDIALVQLEKPSNLTETIKFADSDKLRVGDFTVAIGNPFGIQGVTSIGIVSALGRSTGS 203
QY 229 ES--YTPRIQTDVAINPNSGGPFLNPKGOVVGINSQIYRSRSGFMGTSFATPIDVAMV 286
DB 204 DSGTYENYIOTDAANRNGSGALVNLNGLIGINTAIISSPSGNGAIGAFAPSNQASNL 263
QY 287 AEQLKNTGKVGQGLGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIY 346
DB 264 VQQLIERGQVRGLIGIKGELNADLAKAFNVSAQGAFFSEVLPKSAAEKAGLKAAGDII 323
QY 347 LSLDGEIRSSGDLPMVVGAIITPGKEVSLGVWRKGEETITPAKLGNAAEHTGASSTDEA 406
DB 324 TAMNGOKISSFAIRAKIATGAGKEISLTVLRDGSMDVKKML-QADSSQLSSTKT-EL 381
QY 407 PYTEQSGTFSVSAGITLTQHTDSSGKHLVVVAV--SDAARAGLRHGDILLAV 459
DB 382 PALD-----GATLKDYDAKGVKGIETIKTOPNSLAQR-GIKSGDIIIGI 425

RESULT 15
US-08-487-167-2
Sequence 2, Application US/08487167
Patent No. 5863302
GENERAL INFORMATION:
APPLICANT: LOOSMORE, Sheena M.

APPLICANT: YANG, Yan-Ping
APPLICANT: CHONG, Pele
APPLICANT: COMEN, Raymond P.
APPLICANT: KLEIN, Michel H.
TITLE OF INVENTION: Analog of Haemophilus Hinf7 Protein with
REDUCED PROTEASE ACTIVITY
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sim & Mcburney
STREET: Suite 701, 330 University Avenue
CITY: Toronto
STATE: Ontario
COUNTRY: Canada
ZIP: M5G 1R7
COMPUTER READABLE FORM:
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/487,167
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/296,149
FILING DATE: 26-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/278,091
FILING DATE: 21-JUL-1994
ATTORNEY/AGENT INFORMATION:
NAME: Stewart, Michael I.
REGISTRATION NUMBER: 24,973
REFERENCE/DOCKET INFORMATION:
TELECOMMUNICATION INFORMATION:
TELEPHONE: (416) 595-1155
TELEFAX: (416) 595-1163
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 463 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-487-167-2

Query Match 26.9%; Score 632; DB 2; Length 463;
Best Local Similarity 38.5%; Pred. No. 1.3e-50;
Matches 160; Conservative 75; Mismatches 153; Indels 28; Gaps 12;

QY 54 LPDPAQLVQSEGPVAVNIQAAPARTONGSGNAETDS-DPLADSDP--FYEPF-KRLVYN 109
DB 28 LPSPVSEQNSLAPMLEKVQ--PAVVTLSVSGKAKVDSRSPFLDDIPEEFKFFFGDRFAEQ 85
QY 110 MPEIPOEADDDGLNFGSGFII-SKNGYILTNTHVAVGMGSIKYLLNDKREYTAKLIGSD 168
DB 86 FGGRGESKRNFRGL--GSGVILNASKGYVLTNNHVVIDEADKITVQLODGRERAKLVGKD 143
QY 169 VQSDVALLKIDATEELPVVKIKNPKNLKPGEWAAIGAPFFGDNSTYAGIVSAKRSPLN 228
DB 144 ELSDIALVQLEKPSNLTETIKFADSDKLRVGDFTVAIGNPFGIQGVTSIGIVSALGRSTGS 203
QY 229 ES--YTPRIQTDVAINPNSGGPFLNPKGOVVGINSQIYRSRSGFMGTSFATPIDVAMV 286
DB 204 DSGTYENYIOTDAANRNGSGALVNLNGLIGINTAIISSPSGNGAIGAFAPSNQASNL 263
QY 287 AEQLKNTGKVGQGLGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPAERAGLQAGDIY 346
DB 264 VQQLIERGQVRGLIGIKGELNADLAKAFNVSAQGAFFSEVLPKSAAEKAGLKAAGDII 323
QY 347 LSLDGEIRSSGDLPMVVGAIITPGKEVSLGVWRKGEETITPAKLGNAAEHTGASSTDEA 406
DB 324 TAMNGOKISSFAIRAKIATGAGKEISLTVLRDGSMDVKKML-QADSSQLSSTKT-EL 381
QY 407 PYTEQSGTFSVSAGITLTQHTDSSGKHLVVVAV--SDAARAGLRHGDILLAV 459

Db 382 PALD-----GATLKDYDAKGVKGIEITRKIQPNSLAAQR-GLKSGDIIIGI 425

Search completed: July 11, 2003, 10:34:33
Job time : 29 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:33:18 ; Search time 54 Seconds
(without alignments)
1002.504 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKKYQFALALCALIAG.....ERAGLRHDELLAVRASPQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 445758 segs, 116419773 residues
1 number of hits satisfying chosen parameters: 445758

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA.*
1: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2234.5	95.0	498	10 US-09-388-089B-11	Sequence 11, Appl
2	2130.5	90.5	475	10 US-09-388-089B-12	Sequence 12, Appl
3	1992.5	84.7	447	10 US-09-388-089B-2	Sequence 2, Appl
4	990.5	42.1	460	9 US-09-975-719-132	Sequence 132, App
5	731.5	31.1	503	10 US-09-752-385-8	Sequence 8, Appl
6	705	30.0	474	9 US-10-210-296-7	Sequence 7, Appl
7	559.5	23.8	387	9 US-09-895-913A-120	Sequence 120, App
8	476.5	20.3	619	9 US-10-156-761-10652	Sequence 10652, A
9	455.5	19.4	472	9 US-10-156-761-11757	Sequence 11757, A
10	453	19.3	549	9 US-09-712-363-190	Sequence 190, App
11	443.5	18.8	403	9 US-10-102-806-552	Sequence 552, App
12	442.5	18.8	413	9 US-09-738-626-4478	Sequence 4478, App
13	418	17.8	464	9 US-09-712-363-182	Sequence 182, App
14	414.5	17.6	458	9 US-10-197-634-1	Sequence 1, Appl
15	414	17.6	348	9 US-09-796-753-34	Sequence 34, Appl
16	412.5	17.5	453	9 US-09-796-753-32	Sequence 32, Appl
17	412.5	17.0	476	10 US-09-935-390A-37	Sequence 37, Appl
18	399.5	15.2	286	10 US-09-764-898-256	Sequence 256, App
19	356.5	15.2	411	9 US-10-156-761-11579	Sequence 11579, A

20	348	14.8	330	10 US-09-764-898-184	Sequence 184, App
21	313.5	13.3	178	9 US-09-969-384-18	Sequence 18, Appl
22	312	13.3	9	US-09-712-363-161	Sequence 161, Appl
23	285.5	12.1	255	9 US-09-866-050A-694	Sequence 694, Appl
24	242	10.3	51	10 US-09-388-089B-16	Sequence 16, Appl
25	213.5	9.1	596	10 US-09-287-849-26	Sequence 26, Appl
26	213.5	9.1	729	10 US-09-287-849-2	Sequence 2, Appl
27	206	8.8	361	9 US-10-197-634-15	Sequence 15, Appl
28	184	7.8	397	9 US-09-712-363-280	Sequence 1280, App
29	156	6.6	394	9 US-10-156-761-13542	Sequence 13542, A
30	141	6.0	399	9 US-10-156-761-12129	Sequence 12129, A
31	137.5	5.8	767	10 US-09-219-497-59	Sequence 59, Appl
32	136.5	5.8	452	9 US-10-156-761-15025	Sequence 15025, A
33	136	5.8	2037	9 US-09-951-402-3	Sequence 3, Appl
34	136	5.8	2037	10 US-09-951-402-3	Sequence 3, Appl
35	136	5.8	2037	10 US-09-922-101-3	Sequence 3, Appl
36	131	5.6	68	10 US-09-864-761-34281	Sequence 34281, A
37	130	5.5	724	9 US-10-211-962-21	Sequence 21, Appl
38	129.5	5.5	397	9 US-09-738-626-3831	Sequence 3831, Ap
39	128.5	5.5	204	10 US-09-925-300-1397	Sequence 1397, Ap
40	128	5.4	267	10 US-09-735-705-352	Sequence 352, App
41	128	5.4	267	10 US-09-850-716A-352	Sequence 352, App
42	128	5.4	267	10 US-09-897-778-352	Sequence 352, App
43	128	5.4	683	10 US-09-841-132-357	Sequence 357, App
44	126.5	5.4	434	9 US-10-156-761-10099	Sequence 10099, A
45	125	5.3	304	9 US-10-012-896-835	Sequence 835, App

ALIGNMENTS

RESULT 1																					
US-09-388-089B-11																					
; Sequence 11, Application US/09388089B																					
; Patent No. US20020018782A1																					
; GENERAL INFORMATION:																					
; APPLICANT: Jackson, W.																					
; APPLICANT: Harris, A.																					
; TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID																					
; TITLE OF INVENTION: SEQUENCE AND USES THEREOF																					
; FILE REFERENCE: 7969-083																					
; CURRENT APPLICATION NUMBER: US/09/388,089B																					
; CURRENT FILING DATE: 1999-08-31																					
; NUMBER OF SEQ ID NOS: 20																					
; SOFTWARE: PatentIn Ver. 2.0																					
; SEQ ID NO 11																					
; LENGTH: 498																					
; TYPE: PRT																					
; ORGANISM: Neisseria meningitidis																					
US-09-388-089B-11																					
Query Match																					
Best Local Similarity 95.0%; Score 2234.5; DB 10; Length 498;																					
Matches 444; Conservative 5; Mismatches 13; Indels 1; Gaps 1																					
QY	1	VF	KKYQF	ALAL	CA	LLAG	CEK	AGS	FRG	ADK	KEAS	FVER	IET	HK	DDG	SV	ML	DP	FQ	L	
DB	1	VF	KKYQY	LAL	AL	CA	AS	LAG	CD	KAS	FF	GA	KKEAS	FVER	I	KHT	DDG	SV	ML	DP	FQ
QY	61	VO	SE	GP	AV	NI	QA	PA	PT	ON	GS	GN	AE	TD	SD	PL	AD	SP	FE	FF	
DB	61	VO	SE	GP	AV	NI	QA	PA	PT	ON	GS	NA	ET	SD	PL	AD	SD	PF	FE	FF	
QY	121	GG	IN	FG	SG	FT	IS	KG	YL	IN	TH	VA	MG	SI	KV	LI	ND	RE	Y	AK	
DB	121	GG	IN	FG	SG	FT	IS	KD	GY	IL	IN	TH	VV	TG	MS	IK	VL	ND	KRE	Y	
QY	181	TE	EL	PV	YK	IG	NP	KL	KG	EV	AA	IG	AP	FG	DN	ST	AG	IV	AK	GS	
DB	181	TE	EL	PV	YK	IG	NP	KL	KG	EV	AA	IG	AP	FG	DN	ST	AG	IV	AK	GS	
QY	241	IN	FG	SG	FT	IS	KG	YL	IN	TH	VA	MG	SI	KV	LI	ND	RE	Y	AK	GS	
DB	241	IN	FG	SG	FT	IS	KG	YL	IN	TH	VA	MG	SI	KV	LI	ND	RE	Y	AK	GS	

Db 240 INFGNSGGLFNLKQGVVINSQIYRSRSGFMSGFIPIDIVANVNAEOLKNTKRVORQ 299
Qy 301 LGVITIEVSYGLAOSFGLDKASGALLIAKILPGSPAERAGLOAGDIYLSLDGEIRSSGDL 360
Db 300 LGVITIEVSYGLAOSFGLDKASGALLIAKILPGSPAERAGLRAGIIVLSLDGEIRSSDL 359
Qy 361 PVMVGAITPKEKESVSLGVMRKGEETITIKLGNAAEHTGASKTDEAPYTEQSGTFVSES 420
Db 360 PVMVGAITPKEKESVSLGVMRKGEETITIKLGNAAEHTGASKTDEAPYTEQSGTFVSES 419
Qy 421 AGTTLQHTDSSGKHLVVRVSDAERAGLRGDEILAVRASP 463
Db 420 AGTTLQHTDSSGKHLVVRVSDAERAGLRGDEILAVQGP 462

RESULT 2

US-09-388-089B-12
; Sequence 12, Application US/09388089B
; Patent No. US20020018782A1

GENERAL INFORMATION:
APPLICANT: Jackson, W.

APPLICANT: Harris, A.

TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID

FILE REFERENCE: 7969-083

CURRENT APPLICATION NUMBER: US/09/388, 089B

NUMBER OF SEQ ID NOS: 20

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 12

LENGTH: 475

TYPE: PRT

ORGANISM: Neisseria meningitidis

US-09-388-089B-12

Query Match

Best Local Similarity 90.5%; Score 2130.5; DB 10; Length 475;

Matches 424; Conservative 4; Mismatches 11; Indels 1; Gaps 1;

Qy 24 ACSFPGADKKEASFVRIEHTKDDGSVMLLPFPALQVSEGPVAVNIQAAPARTONGS 83
Db 1 ACSFPGADKKEASFVRIEHTKDDGSVMLLPFPALQVSEGPVAVNIQAAPARTONGS 60
Qy 84 GNAETDSDPLADSDPFYEFKRLVPMNPBIPOEADDDGILNFGSGFIISKNGYILTNTHV 143
Db 61 SNAETDSDPLADSDPFYEFKRLVPMNPBIPOEADDDGILNFGSGFIISKNGYILTNTHV 120
Qy 144 VGMGSIKYLNDKREYTAQLIGSDVQSDVALLKIDATEELPVVKIGNPKLKEGVAA 203
Db 121 VGMGSIKYLNDKREYTAQLIGSDVQSDVALLKIDATEELPVVKIGNPKLKEGVAA 180
Qy 204 IGAPFGDNSVTAGIYSAKGRSLPNESYTPFIOTDVAINFGNSGGLFNLKQGVVINSQ 263
Db 181 IGAPFGDNSVTAGIYSAKGRSLPNESYTPFIOTDVAINFGNSGGLFNLKQGVVINSQ 239
Qy 264 IYRSRSGFMSGFIPIDIVANVNAEOLKNTKRVORQGLVYIIEVSYGLAOSFGLDKASG 323
Db 240 IYRSRSGFMSGFIPIDIVANVNAEOLKNTKRVORQGLVYIIEVSYGLAOSFGLDKASG 299
Qy 324 ALIAKILPSSPARAGLOAGDIYLSLDGEIRSSGDLPMVVGAIITPKEKESVSLGVMRKGE 383
Db 300 ALIAKILPSSPARAGLRAGDIYLSLDGEIRSSGDLPMVVGAIITPKEKESVSLGVMRKGE 359
Qy 384 ITIKAKLGNAAEHTGASKTDEAPYTEQSGTFVSESAGITLQHTDSSGKHLVVRVSD 443
Db 360 ITIKAKLGNAAEHTGASKTDEAPYTEQSGTFVSESAGITLQHTDSSGKHLVVRVSD 419
Qy 444 AERAGLRGDEILAVRASP 463
Db 420 AERAGLRGDEILAVQGP 439

RESULT 3

US-09-388-089B-2
; Sequence 2, Application US/09388089B
; Patent No. US20020018782A1
; GENERAL INFORMATION:
APPLICANT: Jackson, W.
APPLICANT: Harris, A.
TITLE OF INVENTION: NEISSERIA MENINGITIDIS POLYPEPTIDE, NUCLEIC ACID
FILE REFERENCE: 7969-083
CURRENT APPLICATION NUMBER: US/09/388, 089B
NUMBER OF SEQ ID NOS: 20
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 447
TYPE: PRT
ORGANISM: Neisseria spp.
US-09-388-089B-2

Query Match

Best Local Similarity 84.7%; Score 1992.5; DB 10; Length 447;

Matches 397; Conservative 3; Mismatches 11; Indels 1; Gaps 1;

Qy 52 MLDPFVQLVQSEGPVAVNIQAAPARTONGSNAETSDPLADSDPFYEFKRLVPMNP 111

Db 1 MLDPFVQLVQSEGPVAVNIQAAPARTONGSSNAETSDPLADSDPFYEFKRLVPMNP 60

Qy 112 EIPOEADDDGILNFGSGFIISKNGYILTNTHVAVAMGSIKYLNDKREYTAQLIGSDVS 171

Db 61 EIPOEADDDGILNFGSGFIISKNGYILTNTHVAVAMGSIKYLNDKREYTAQLIGSDVS 120

Qy 172 DVALLKIDATEELPVVKIGNPKLKEGVAAIGAPFGDNSVTAGIYSAKGRSLPNESY 231

Db 121 DVALLKIDATEELPVVKIGNPKLKEGVAAIGAPFGDNSVTAGIYSAKGRSLPNESY 179

Qy 232 TPIQTDVAINFGNSGGLFNLKQGVVINSQIYRSRSGFMSGFIPIDIVANVNAEOLK 291

Db 180 TPIQTDVAINFGNSGGLFNLKQGVVINSQIYRSRSGFMSGFIPIDIVANVNAEOLK 239

Qy 292 NTGVORQGLVYIIEVSYGLAOSFGLDKASGALLIAKILPGSPAERAGLOAGDIYLSLDG 351

Db 240 NTGVORQGLVYIIEVSYGLAOSFGLDKASGALLIAKILPGSPAERAGLOAGDIYLSLDG 299

Qy 352 GEIRSSGDLPMVVGAIITPKEKESVSLGVMRKGEETITIKLGNAAEHTGASKTDEAPYTEQ 411

Db 300 GEIRSSGDLPMVVGAIITPKEKESVSLGVMRKGEETITIKLGNAAEHTGASKTDEAPYTEQ 359

Qy 412 QSGTFVSESAGITLQHTDSSGKHLVVRVSDAERAGLRGDEILAVRASP 463

Db 360 QSGTFVSESAGITLQHTDSSGKHLVVRVSDAERAGLRGDEILAVQGP 411

RESULT 4

US-09-975-719-132

; Sequence 132, Application US/09975719

; Publication No. US20030022349A1

GENERAL INFORMATION:

APPLICANT: Ausubel, Frederick M.

APPLICANT: Rahme, Laurence G.

TITLE OF INVENTION: VIRULENCE-ASSOCIATED NUCLEIC ACID

FILE REFERENCE: 00786/361003

CURRENT APPLICATION NUMBER: US/09/975, 719

PRIOR FILING DATE: 2001-10-10

PRIOR APPLICATION NUMBER: US 09/199,637

PRIOR FILING DATE: 1998-11-25

PRIOR APPLICATION NUMBER: US 60/066,517

NUMBER OF SEQ ID NOS: 437

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 132

LENGTH: 460

TYPE: PRT

ORGANISM: Pseudomonas aeruginosa
US-09-975-719-132

Query Match 42.1%; Score 990.5; DB 9; Length 460;

Best Local Similarity 52.5%; Pred. No. 1.3e-71;
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

54 LPDPAQLVQSEGPVAVNI---QAAPARTONGSNAETSDPLADDP-FYEFKRLVPM 109
27 LPDFTPLVQSPAVVNISTRQLP---DRAMARQQLSPDLGLPMPMRDLERTIPQ 82
110 MPEIR---QEEADDDGLNFGSGFIIKNGYILTNTHVAVGMSIKYLANDKREYTA 166
83 VPRNRGQGRQAQ---SLGSGFIIISNDGYILTNHVVADADAILVRLSRSHKAKLVG 138
167 SDVQSDVALLKTDATTEELPVVKIGNPKNLKPGEWVAATGAPRGFDSVYTAGIYSA 226
139 ADPRSDVAVLKEA-KNLPFLKLGDSNKLKVGEMVVAITGSPFGFDSVYTAGIYSA 197
227 PNESTYPTIQTVAINPGNSGGPLFNLKGVVQVINSQIYSRSGFMGISFAIPIDVAM 286
198 PNESTYPTIQTVAINPGNSGGPLNLEGEVQVINSQIIFRSFGFGLSFAIPIDVAM 257
287 ABOLKNTKVGORGVITIIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAG 346
258 ADQLKAGKVRSGWLGVIQEVNKLDAESFGLDKPSGALVAVQLVEDGPAKGLQVGD 317
347 LSLDGEIRSSGDLPMVNAITPGEKVSIGVWRKGEITIKALGNAAHTTGASKTDE 405
318 LSLNQSINESADLPPLVGNMKRGKINLDVIRNGRKSLSMAVGNLPD-----DDEE 370
406 ---APYTEOQSGTFVSAGITLQTH--TDSSGKHLVVVRVSDA-AERAGLRHGEI 456
371 IASMAFPAERSNNLGVTVADLTAEQRKSLDQG--GVYIKVQDGPAAVYIGLRPGDV 428

RESULT 5

US-09-752-385-8
Sequence 8, Application us/09752385
Patent No. US20020012919A1

GENERAL INFORMATION:

APPLICANT: Anderson, Burt E.

Regnery, Russell L.

TITLE OF INVENTION: Nucleic Acids of Rochalimaea Henselae

and Methods and Compositions for Diagnosing

Rochalimaea

Henselae and Rochalimaea Quintana Infection

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: NEEDLE & ROSENBERG, P.C.

STREET: 127 Peachtree Street, N.E., Suite 1200

CITY: Atlanta

STATE: Georgia

COUNTRY: USA

ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/752,385

FILING DATE: 29-Dec-2000

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 09/525,310

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Sopart, Gwendolyn D.

REGISTRATION NUMBER: 36,016

REFERENCE/DOCKET NUMBER: 1414.624

TELECOMMUNICATION INFORMATION:

TELEPHONE: (404) 688-0770

TELEFAX: (404) 688-9880

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:
LENGTH: 503 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein

SEQUENCE DESCRIPTION: SEQ ID NO: 8:

US-09-752-385-8
Query Match 31.1%; Score 731.5; DB 10; Length 503;
Best Local Similarity 39.5%; Pred. No. 1e-50;
Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;

8 FALAALCAALLAGEKAGSPFGADKKEASFVERIEHTDDGSVSLPLDPAQLVQSEGA 67
15 FSALELTALFFSGC--GSSLMTTKAHNSV-----FSSLMOQGFADIVQVQKPA 62
68 VVNIQAAPARTON---GSGNAETSDPLADSDP---FYEFKRLVPMPEIPOEADD 120
63 VVSQVQVSNKKKKKWFSPDFSTPGPQLPDQHPLKRFQDFVNRDKFSNKL-QRSHRL 121
121 GGLNFGSGFIIISKNGYILTNTHVAVGMSIVYLANDKREYTAKLIGSDVQSVALLK 180
122 RPIAFGSGFPISSDGYIVTNHVISDGTSYAVVLDGTELNAKILGTDPRDLAVLKVE 181
181 TEBLPVVKIGNPKNLKPGEWVAATGAPRGFDSVYTAGIYSAKGRSLPNESTYPTIQT 240
182 KRKFSYVDFGDSKLRAGDVVAAGNPGGLGTVYAGIYSAKGRDITGVYDDFIQIDA 241
241 INPGNSGGPLFNLKGVVQVINSQIYSRSGFMGISFAIPIDVAMVNAEOLKNTGVORQ 300
242 VNKNSSGPFITDLNGKVVGVNTAIFSPSGGVGIAFAIPATFANVQQLLEKGLVGRW 301
301 LGVITIGEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLQAGDIVSLDGEIRSS 360
302 LGVQIQVTKRISDSIGLKEAKGALITDPLKG-PAKAKIRAGDVIIISVNGEKINDV 360
361 PVMGATTPGKEVSLGVWRKGEITIKAKGNAAHTTGASKTDEAPYT-EGQSGTFVS 419
361 AKRIANNSPGETVVLGVWKSKEENIKVLDSPMD--ENMKDSKYSNENGSDETLE 417
420 SAGITLQTHDSSGKHLVVVRV---SDAERAGLRHGEIILAV 459
418 DYGLIAPSDDGLG--LVTVDPEDSDADK-GIRPGDVITV 457

RESULT 6

US-10-210-296-7

Sequence 7, Application US/10210296

Publication No. US20030021802A1

GENERAL INFORMATION:

APPLICANT: Pfizer Products Inc.

TITLE OF INVENTION: LAWSONIA INTRACELLULARIS PROTEINS, AND RELATED

TITLE OF INVENTION: METHODS

TITLE OF INVENTION: AND MATERIALS

FILE REFERENCE: PC10589A

CURRENT APPLICATION NUMBER: US/10/210,296

CURRENT FILING DATE: 2002-08-01

PRIOR APPLICATION NUMBER: US/09/689,065

PRIOR FILING DATE: 2000-10-12

NUMBER OF SEQ ID NOS: 102

SOFTWARE: Patent in Ver. 2.1

SEQ ID NO 7

LENGTH: 474

TYPE: PRT

ORGANISM: Lawsonia intracellularis

US-10-210-296-7

Query Match 30.0%; Score 705; DB 9; Length 474;

Best Local Similarity 39.3%; Pred. No. 1.3e-48;

Matches 169; Conservative 67; Mismatches 152; Indels 42; Gaps 11;

QY	54	PDFFQOLVQSEBPAAVNIQA-APAPRTQNGSGNAETDSDPLADSDPEYE-FPRKLVPMNP	1111
Dd	28	LPNFPLPKDASKAVVNIISTEKKIPR-----GRTEPFMEFPRGLPROFENFEQFEKCP	82
QY	112	--EIPQSEADOGGLNFGSGFIISKNGYILTNTHVAVGMSIKVLL-----NDKREYTAKL	1655
Dd	83	DSQIHKKR-----SLGTGFIISSDDYITNNHVIEGADSVRNLEGTSGKESSLPRAVI	1366
QY	166	GSDVQSDVALLKIDATTEELPVYKIGRPNKULPQEWAAIACA PRGFQDVSYAGIVSAAGRS	225
Dd	137	GRDESTDALLKVKAKSDSLPYLLFGNSDMEVGEWALAGNPRGLGHTVYAGILLSAAGR	196
QY	226	LPNESYTFPIQTDVAINPFGNSGGPLFNLGQVAVNGVINSQIYSRSGGFMGISFPAIPIDVAMN	2855
Dd	197	IHAGFBNDFLOTDASINPENSNGPFLNMGQVVGINTALMA-SG--QGIGFALPSSMAHR	255
QY	286	VAEQUKATGKVRQGLGVITQEVSYGCLAQSGFLGDKASGALLIAKILPGSPAPRAGLOAGDT	345
Dd	254	IIEQLKTKKYSRGMIGVITQDVDTNTAALGLSQAKGALVGSVVPQDPAADKGLKVGDI	3133
Dd	346	VLSLDGGEIRSSGDLPVWVGATITPCKEVEVLGWRKCEBITIKAKLGNALNHTGASSKTE	405
Dd	314	VTQADGKQIDASSLKAIATATPRPSVVLKWKWRDGSKDITSLGERRKTTSSQKQSSPE	3737
QY	406	A-----PYTEQSGTFPSVESAGITLQTHTSSGKRLVVVRVSDAERAAGLRHGE	4555
Dd	374	SLPGALGSLVRPLTQESKSFQVK-LGI-----GLLVEVEPNKPSAEGAIHQDI	4233
QY	456	ILAVYASPRQ	465
Dd	424	ILSANLKPRQ	433

```

RESULT 7
US-09-895-913A-120
; Sequence 120, Application US/09895913A
; Patent No. US20020160456A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020160456A1e1 Helicobacter Polypeptides in the
; TITLE OF INVENTION: Genome
; FILE REFERENCE: 06132/043002
; CURRENT APPLICATION NUMBER: US/09/895,913A
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 08/861,227
; PRIOR FILING DATE: 1997-06-24
; NUMBER OF SEQ ID NOS: 368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 120
; LENGTH: 387
; TYPE: PRT
; ORGANISM: Helicobacter pylori
; US-09-895-913A-120

```

	Query Match	23.8%	Score 559.5:	DB 9;	Length 387;
	Best Local Similarity	36.9%*	Pred. No. Se-37;		
	Matches 131; Conservative	73;	Mismatches 134;	Indels 17;	Gaps 7.
Oy	113 I P O E A D D G C L N F G S G F T I I S K R G Y I L T H V W A G M S I K V L L - N D K R E T A L I G S D V O S	171	:	: :	:
Db	2 I P K R M E R B A -- L G S G V I I S X D G Y I V T N H V I D G A D K I K V T I P G S N K E Y S A T L V G D N S E S	58	:	: : :	:
Oy	172 D V A L L K I D A T E L P V V K I G N P R U L K P G E W A A I G A P F P D N S V T A G I Y S A K R S - L P N E S	230	:	: :	:
Dd	59 D L A V I R I - T K D N L P T I K F S D S N D I S V G D L V F A I G N P F G S E S V T O G I A L N K S G I G I N S	117	:	: : :	:
Oy	231 Y T P I Q I D V A I N P G N S G P L R N L K Q V V G I N S Q I S B R G G F M G I S F A L P I D V A N V A E L	290	:	: : :	:

```

Db      118 YENFIQTDA$INPGNSGALLIDRGLVINGINIAIISKTGNGNHGIGALPISNMWVDIVTQ 177
Oy      291 KNTGKVGORGLGVITIQEVSYGLAOSFGILDKASGALLIAKIPGSPARAGLQAGDIYLSLD 350
        |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      178 IKTGKIEBGVGLGVQLDLSGDLQNSY--DNKEGAVVIVSEKDSPAKKGILLWMDLITEYV 235
Oy      351 GGEIRSSDDLVMWCAITPGKEVSLGVWRKBEIITIKALGNAAEHTGASSITYDEAPYB 410
        |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      236 GKRYKYNTELNLIGSMLPNQRVYTLKVRDKKERAPFTYLL--AEKQNKKEITISAONG 292
Oy      411 QOSGTFVS$AGITLOT-----HTDS$GKHLVVVR$VSDPAERAGLRHDEITLAV 459
        |||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db      293 AOGQNLGQVEDLIVQETGK$MRSLSDVQGVLV$QVNM$SPAQ$GFRQ$NITTKI 347

```

```

RESULT 8
US-10-156-761-10652
; Sequence 10652, Application US/10156761
; Publication No. US20030119018X1
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATTORI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; CURRENT FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 10652
; LENGTH: 619
; TYPE: PR1
; ORGANISM: Streptomyces avermitilis
US-10-156-761-10652

```

Query Match	Best Local Similarity	20.3%	Score 476.5	DB 9	Length 619
Matches 133	Conservative	32.0%	Pred. No. 4.7e-30	Mismatches 142	Indels 77
					Gaps 10
Qy	4	KYQYFALAAALCAALLACCEKAGSEFFGADKKEASFVERIEHTKDDGSVMLLPDFAQIVQS	63		
Db	254	RVRLVLTGAVTLAVLVSG-----GIGAVGVYLER-----TGVTVDVLELPQAQO--ES	298		
Qy	64	EGPRAVNI-----QAAPAPRTQNSGNAETDSDPLASDPYEFEFKULVPMPEIPOEHA	118		
Db	299	EERAAADSVAGIAAALPSVTLTHVKGSAAECT-----	330		
Qy	119	DDGGLNFGSGEIIISKNGYIILTNTHVAVAGMS---IKYLLDNKREYTAKLISDVQSDVAL	175		
Db	331	-----GTGVLVDGRHIIILTNHNVVAPASSGSEISVTFSGEGFAKXKATVVERDSDYDLAV	383		
Qy	176	LKIDATEELPVYKIGNPFNLKPGEWVAALGAPSGFDNSVTAGIYSAKRSI-----	226		
Db	384	VKYSVGVLGKPMPLGNSDNDVQGPVVAALGAPFLNATVTVSGIISAKERPITTAGGEEKGD	443		
Qy	227	PNESYTFPIOTDVAINPNSGGPLFNLKGOVVGINSOIYS-----RSG--GFMGISPAI	278		
Db	444	SDVSYVDALOTDAPINPNSGGPLDLSKARVVGINSAIRADSSDDSGQAGSITGLGPAI	503		
Qy	279	PIDVAMNVAAEOLKNTGFVYORGOGLVIIIOEVSYGLAOSFGLDKASGALITAKLIPSPAREA	338		
Db	504	PVNGAKRVAEELINTGRATHPHVIGVTLDMDYTDGGAUVGTKSDNGG--SPYTRGSGPORA	561		
Qy	339	GLQAGDIVLSIDGGEIRSSGDLPMYVGAITIPGKEVSIQVWRKGEITITTKAKLGNAA	394		
Db	562	GIQAGDVITTEVDGGRINSGBELIYKRIAHNPGRLATLVEEDGKEKPTVLVLSGAS	617		

```
RESULT 9
US-10-156-761-11757
; Sequence 11757, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
; APPLICANT: OMIURA, SATOSHI
; APPLICANT: IKEDA, HARUO
; APPLICANT: ISHIKAWA, JUN
; APPLICANT: HORIKAWA, HIROSHI
; APPLICANT: SHIBA, TADAYOSHI
; APPLICANT: SAKAKI, YOSHIYUKI
; APPLICANT: HATOKI, MASAHIRA
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
; FILE REFERENCE: 249-262
; CURRENT APPLICATION NUMBER: US/10/156,761
; PRIOR FILING DATE: 2002-05-29
; PRIOR APPLICATION NUMBER: JP 2001-204089
; PRIOR FILING DATE: 2001-05-30
; PRIOR APPLICATION NUMBER: JP 2001-272697
; PRIOR FILING DATE: 2001-08-02
; NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 11757
; LENGTH: 472
; TYPE: PRF
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11757

Query Match          19.4%; Score 455.5; DB 9; Length 472;
Best Local Similarity 32.5%; Pred. No. 1.6e-28;
Matches 138; Conservative 53; Mismatches 140; Indels 93; Gaps 14;

QY 11 AALCALLAGCEKASFFGADKKEASFVERIETHTKDDSVSMILPDAQLVSEGPAYVN 70
DB 108 AILVPAALVAGVGVGIGITLAK-----DNQSSGSTTYSASDSGSGVGRDGTAG 158
QY 71 IQAAPART-----QNGSNAETSDPLADSDPFYEFKRLVPMPEIQOEADGGLNF 125
DB 159 VAKALPSTYTIENQSSSG-----EGGT-- 181
QY 126 GSGFIISKNGGILNTNTVAVAGM-----GSIKVLNDKREYAKLIGSDVQSDVALKI-DAT 181
DB 182 GIGFVFPKQGHIVTNHVNVAEVDGKLTATFPDGKKNNAEVGAQGVDAVAVLTKAP 241
QY 182 EELPVVKIGNPKNLKPGEWVAIGAIPFGDINSVTAGIVSAK-----GRELPMESYTP 233
DB 242 SDIQPLTLDGSDKVAVDSTIALIGAPFGLSTVTITGIIISAKRPFVASSDSSSKASTWS 301
QY 234 FIOTDVAINPNSGGPLFNLKGOVVGINSQIYSRS-----GFGMISPAIPIDVAM 284
DB 302 ALQTDASINPNSGGPLLDAGSVIGINSALQSSSSGGIGSSGSGSGLGPAIPINQAK 361
QY 285 NVAEOLKNTKVRQGOGLVIOEVSYGLAQSGFLDK-----ASGALLAKILPGSPAERAG 339
DB 362 YVAOELKTKRPYPVVG-----ASVSLREGTGAKITGOGASGS--DAITPNGPAAAG 414
QY 340 LQAGDIYLSLDGEIRSSGDLPMVWGAIT---TPGEVSLGVWRKGEETITIAKLGNAEH 396
DB 415 LKRGDVTITKDDWIDSG---PTLIGELWTRPQATVATLTTRODKANTTTVTILG---ER 468
QY 397 TCAS 400
DB 469 VGDS 472

RESULT 10
US-09-712-363-190
; Sequence 190, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: EISENBERG, DAVID
; APPLICANT: ROTSTEIN, SERGIO H.
```

```
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; TITLE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 190
; LENGTH: 549
; TYPE: PRF
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-190

Query Match          19.3%; Score 453; DB 9; Length 549;
Best Local Similarity 30.3%; Pred. No. 3.1e-28;
Matches 122; Conservative 68; Mismatches 142; Indels 70; Gaps 10;
```

```
QY 4 KYQYFALAALCA-ALLAGEKASFFGADKKEASFVERIETHTKDDSVSMILPDAQLVQ 62
DB 196 KSVYALAGILVALAVIG--GIGGVIG--RKTAEVVDAFPTSK----- 234
QY 63 SEGPVAVNIQAAPARTQNGSNAETSDPLADSDPFYEFK--RLVPMPEIQOEADD 120
DB 235 -----VLTSTTGNQAQ-----EPAGRETKVAAAVADSVTITESVDQ 270
QY 121 GGLNFGSGFIISKNGGILNTNTVAVAGM-----IKVLNDKREYAKLIGSDVQSDVA 174
DB 271 EGMQ--GSGVTVDRGRTIVTNHVNHVISPAANNPSQFTTTFVFNDEKEVPANLVGRDPKTDLA 329
QY 175 LTKIDATEBELPVVKIGNPKNLKPGEWVAIGAIPFGDINSVTAGIVSAKGRSLP----- 227
DB 330 VLKVNVDNLTVARLGDSSKRVAGDEVLAVGAPLGRSTVTOGIVSALHHPVPLSGGSD 389
QY 228 NESYTPFIOTDVAINPNSGGPLFNLKGOVVGINSQIYSRSGGFMGISFAIPIDVAMVA 287
DB 390 TDTVIDALQTDASINPNSGGPLLDMDAQVIGINTAGKSLSDSASGIGFAIPVNMELVA 449
QY 288 EQLKNTKVRQGOGLVIOEVSYGLAQSGFLDKASGALLAKILPGSPAERAGIQAQGIYL 347
DB 450 NSLIDGKIHPFTLIGSTRSVSNAL-----ASAQAVANVAGSPAQGGIILENDIV 501
QY 348 SLUDGEIRSSGDLPMVWGAITPGEVSLGVWRKGEETITIAK 389
DB 502 KVNRAVADSDDEFVAVNRQLAIGQDAPIVVRAGRHVTILVK 543

RESULT 11
US-10-102-806-552
; Sequence 552, Application US/10102806
; Publication No. US20030054421A1
; GENERAL INFORMATION:
; APPLICANT: ROSEN ET AL.
; TITLE OF INVENTION: NUCLEIC ACIDS, PROTEINS AND ANTIBODIES
; FILE REFERENCE: PA103PIC1
```

```

: LENGTH: 413
: TYPE: PRT
: ORGANISM: Corynebacterium glutamicum
US-09-738-626-4478

Query Match
Best local Similarity 32.8%; Score 442.5; DB 9; Length 413;
Matches 129; Conservative 57; Mismatches 137; Indels 69; Gaps 10;

QY 3 KKYQYFALALCAALLAGCEKAGSFFGADKKEASFYERIEHTYKDDGSVMLLPDPALVQ 62
DB 56 KKKRVGIGFALMLMLVGSFATGSVGV-----AATQLG 89
QY 63 SEGPAAVNIQAAR-PTQNGS-GNAETSDPLADSPFYEFKRLVPMNPEI PQEAD 120
DB 90 SDSSTPNALEQPSVORTTAAEPGSAEQVAVALPS-----VSIQALTRTSASE 139
QY 121 GGLNFGSGFTIISKNGYLLTHTHVAGM--GSIKVLNDKREYTAKLIGSDVQSDVALLK 177
DB 140 -----GSGSIISSDGYVMNNHVAAGIEQGVLEFVSSDGTAAQADTIAADPSTIDIAVIK 194
QY 178 IDATEELPVYKIGNPKNLKPEGVAVALGAPFGFDSVTAGIVASAKRSL-----PNESY 231
DB 195 ITRVSNLPVWMSFGDSPLAGVGQSVMAVGSPLGISSTVTGTGIVASVNRPVASGDCGRSL 254
QY 232 TPTPTQDVAINPENSGGPLFNLKGQVVGINSQI-----YRSGGFMGISPAIPDIVAMNV 286
DB 255 IDAIQTDAAINPENSGGPLVMDMGNLGMNMSVIASISTSDSAGSICLGFSPSNFAKRV 314
QY 287 AEOUKNKGKQVROQLGIIIEVSVGLAQSFGLDKA-SGALIAKLIPSPAPERAGLOAGDI 345
DB 315 ADQILSTGQVTPMIGQV-----GTDNSVTAIVIASVQDGGPADGLOPEDI 363
QY 346 VLSLDGGEIRSGDLPVMVGAITPGKEVSLGV 377
DB 364 VTKLNDKRVIDSPLSLIAAVRSHDPGEIVTLTI 395

RESULT 13
US-09-712-363-182
: Sequence 182, Application US/09712363
: Patent No. US20020164588A1
: GENERAL INFORMATION:
: APPLICANT: Eisenberg, David
: APPLICANT: Rotstein, Sergio H.
: APPLICANT: Marcotte, Edward M.
: TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
: FILE OF INVENTION: INTERACTIONS OF PROTEINS BY COMPARATIVE ANALYSIS
: FILE REFERENCE: 07419-032001
: CURRENT APPLICATION NUMBER: US/09/712.363
: CURRENT FILING DATE: 2000-11-13
: PRIOR APPLICATION NUMBER: PCT/US00/02246
: PRIOR FILING DATE: 2000-01-28
: PRIOR APPLICATION NUMBER: 60/179,531
: PRIOR FILING DATE: 2000-02-01
: PRIOR APPLICATION NUMBER: 60/117,844
: PRIOR FILING DATE: 1999-01-29
: PRIOR APPLICATION NUMBER: 60/118,206,
: PRIOR FILING DATE: 1999-02-01
: PRIOR APPLICATION NUMBER: 60/126,593
: PRIOR FILING DATE: 1999-03-26
: PRIOR APPLICATION NUMBER: 60/134,093
: PRIOR FILING DATE: 1999-05-14
: PRIOR APPLICATION NUMBER: 60/134,092
: PRIOR FILING DATE: 1999-05-14
: PRIOR APPLICATION NUMBER: 60/165,124
: PRIOR FILING DATE: 1999-11-12
: PRIOR APPLICATION NUMBER: 60/165,086
: PRIOR FILING DATE: 1999-11-12
: NUMBER OF SEQ ID NOS: 292
: SOFTWARE: PatsSeq for Windows Version 4.0
: SEQ ID NO 182
: LENGTH: 464

```

```

; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-182

Query Match      17.8%; Score 418; DB 9; Length 464;
Best Local Similarity 32.5%; Pred. No. 1.6e-25;
Matches 116; Conservative 60; Mismatches 139; Indels 42; Gaps 10;

Qy 67 AVNNIQAAPAPRTQNGSNAETDSDPLADSDP--FYEFPKALVPMNBEIPOEADOGGLN 124
Db 122 SLVGNRRAPAGSGSGPVAAASAPSPIDAAAMPBGSVQVAAKVPSPVML---ETDGRGS 178
Qy 125 -FGSGPIISKNGYILTNTHVAG-----MGS-----IKVLNDRREYAKLIGSDVSDVA 174
Db 179 EEGSGIILSAEGLILTNHNVIAAAKAPLPGSPPKTTVTFSDRTAPFTVAGADPFSIDIA 238
Qy 175 LKIDATELPPVKIKGNPKLKGEMVAALIGAPFGFDSNVTAGIVAKGRSLP----- 227
Db 239 VVRVQVSGVGLPPIISGSSSDLVVQGPVLAIGSPLGLEGTVTGIVSALNRPVSTTGEAGN 238
Qy 228 NESYPPFIQTDVAIPNGSGGPIFNKGGVNGINSQIYS-----RSGGFMGISFAIPI 280
Db 239 QNTVDAIQTDPAITPGNSGGLVANNADLVGNNSAIAITGADSAQSGSIGLGRPIIV 358
Qy 281 DVANNVAQELKNTGKVRQGLGVIIQEVSYGLAQSFGLDKAS--GALIAKILPGSPAPERAG 339
Db 359 DQAKRIADELISGTAKASHASIGVQVTN-----DKDTLGAKIVEVAVAGAAANAG 407
Qy 340 LQAGDIVSLDGEIRSSGDLFVPMGALTPGKEVSLGVM--RKGEETIRAKLGNAAE 395
Db 408 VPKGVVTVKVDPRPINSADALVAAVRSKAPGATVALTFDPPSGSRTVQTLCKAKQ 464
```

```

RESULT 14
US-10-197-634-1
; Sequence 1, Application US/10197634
; Publication No. US20030073629A1
; GENERAL INFORMATION:
; APPLICANT: Alnemi, Emad S.
; TITLE OF INVENTION: OMT AND DOMAINS THEREOF THAT DISRUPT
; FILE REFERENCE: 480140.479
; CURRENT FILING DATE: 2002-07-15
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 458
; TYPE: PRT
; ORGANISM: Homo sapiens
US-197-634-1

Query Match      17.6%; Score 414.5; DB 9; Length 458;
Best Local Similarity 31.8%; Pred. No. 3e-25;
Matches 110; Conservative 65; Mismatches 108; Indels 63; Gaps 10;

Qy 58 AQLVSEBPAAVNIQAAPARTQNGSNAETDSDPLADSDPFEFPKALVPMNBEIPOE 117
Db 151 ADVVEKTPAAVYIIE-----ILDRHP---FLGREVP----- 178
Qy 118 ADDGGLNFGSGPIISKNGYILTNTHVAGGSIKVLNDRREYAKLIGSDVSDVALLK 177
Db 179 -----ISNGSGVVAADGLIVNAHVADRRAVRVLLSGDITYEAVTVANDPADIATLIR 233
Qy 178 IDATELPPVKIKGNPKLKGEMVAALIGAPFGFDSNVTAGIVS-----AKGRSLPNESSYT 232
Db 234 IQTKPEPLPTLPGRGADVAGQGEFVAVMGSPFALQNTITSGIVBSAQRPAIDLPLP-QTNV 292
Qy 233 PEIQDVAINPNSGGLPFLNKGQVGVINSQIYSRSGFMGISFAIPIDVANNVAE--QL 230
Db 293 EYIQTDPAIDFNSGGLPVLNDEGVGVNMTKYT-----AGISFALPSPDLRPLRGEK 347
Qy 291 KNTGK-----VQRGQLGVIIQEVSYGLAQSFGL-----DKASGALIAKILPGSPAERAGL 340
```

```

Db 348 KNSSSGIGSQRRYIGVMMTLTSPSILAEQLREPSFPDVOHGVLIHKVILIGSPAPRAAGL 407
Qy 341 QAGDIVSLDGEIRSSGDLFVPMGALTPGKEVSLGVMRKGEETIR 386
Db 408 RPDVILAIIGEQWQNAED--VYEAFTQSQALAVQIRGRETLTL 450

RESULT 15
US-09-796-753--34
; Sequence 34, Application US/09796753
; Publication No. US20030027998A1
; GENERAL INFORMATION:
; APPLICANT: McCarthy, Sean A.
; TITLE OF INVENTION: SECRETED PROTEINS AND USES THEREOF
; FILE REFERENCE: 7853-227-999
; CURRENT FILING DATE: US/09/796,753
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 09/183,175
; PRIOR FILING DATE: 1998-10-30
; PRIOR APPLICATION NUMBER: 09/223,094
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/223,546
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/224,246
; PRIOR FILING DATE: 1998-12-30
; PRIOR APPLICATION NUMBER: 09/259,388
; PRIOR FILING DATE: 1999-02-26
; PRIOR APPLICATION NUMBER: 60/122,458
; PRIOR FILING DATE: 1999-03-01
; PRIOR APPLICATION NUMBER: 09/312,359
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 09/336,536
; PRIOR FILING DATE: 1999-06-18
; PRIOR APPLICATION NUMBER: 09/342,687
; PRIOR FILING DATE: 1999-06-29
; PRIOR APPLICATION NUMBER: 09/345,464
; PRIOR FILING DATE: 1999-06-30
; PRIOR APPLICATION NUMBER: 09/365,164
; PRIOR FILING DATE: 1999-07-30
; PRIOR APPLICATION NUMBER: 09/399,723
; PRIOR FILING DATE: 1999-09-20
; PRIOR APPLICATION NUMBER: 09/409,634
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 09/471,179
; PRIOR FILING DATE: 1999-12-23
; PRIOR APPLICATION NUMBER: 09/474,071
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/474,072
; PRIOR FILING DATE: 1999-12-29
; PRIOR APPLICATION NUMBER: 09/514,010
; PRIOR FILING DATE: 2000-02-25
; PRIOR APPLICATION NUMBER: 09/516,745
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 09/572,002
; PRIOR FILING DATE: 2000-05-14
; PRIOR APPLICATION NUMBER: 09/597,993
; PRIOR FILING DATE: 2000-06-19
; PRIOR APPLICATION NUMBER: 09/599,596
; PRIOR FILING DATE: 2000-06-22
; PRIOR APPLICATION NUMBER: 09/630,334
; PRIOR FILING DATE: 2000-07-31
; PRIOR APPLICATION NUMBER: 09/606,565
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/606,317
; PRIOR FILING DATE: 2000-06-29
; PRIOR APPLICATION NUMBER: 09/665,666
; PRIOR FILING DATE: 2000-09-20
; PRIOR APPLICATION NUMBER: 09/677,751
; PRIOR FILING DATE: 2000-09-30
; NUMBER OF SEQ ID NOS: 162
; SEQ ID NO 34
; LENGTH: 348
```

TYPE: PRT
ORGANISM: Mouse
US-09-796-753-34

Query Match 17.6%; Score 414; DB 9; Length 348;
Best Local Similarity 33.7%; Pred. No. 2.3e-25;
Matches 112; Conservative 58; Mismatches 98; Indels 64; Gaps 10;

```
QY 58 AOLVSEGPVAVNIOAPAPRTONGSNAETSDPLADSDPF--YEPFKRLVPMNPEIPIQ 115
DB 39 ADVVEKIAPAVHIEL-----FLRHPLFGKNVP-----66
QY 116 EEADGGILNFGSGFIISKNGYILTNTHVA-----GMSIKVLLNDREXTAKLIGSDV 169
DB 67 -----LSSGSGFIMSEAGLITITNAHVSSNSAPGRQQLKVQLONGDSYEATIKDIDK 119
QY 170 OSDVALLKIDATEELPVVKTGNKPKLPGEWVVAIGAPFGFDSVTAGIVGA---KGRSL 226
DB 120 KSDIATIKIHPKKKLPVLLGHSADLPPGEFVVAIGSPFALQNTVTGTGIVSTAOREGREL 179
QY 227 P-NESTYPIQTDVAINPGNSGGPLFNLKQOVGINSQIVSRSGGFMGISFAIPID-VAM 284
DB 180 GURDSMDYIQTDAIINYNSGGPLVNLDSGVIGINTLKT-----AGISFAIPSDRITR 234
QY 285 NVAE-QLKNTGKVQRQOLGYLIEVSYGLAQSFGL-----DKASGALIAKILPGSPAER 337
DB 235 FLTEPDQKQIKDKMKRPFIGIRMTTITPSLVDELKASNPDPPEVSSGIYVGEVAPNSPSQR 294
QY 338 AGLQAGDIYVLSLDGGEIRSSGDLPMVYGAIIP 369
DB 295 GGIOGDGDIIVKVGRRPLVDSSELQEAVALTESP 326
```

Search completed: July 11, 2003, 10:42:52
Job time : 57 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using SW model

Run on: July 11, 2003, 10:28:37 ; Search time 42 Seconds
(without alignments)
1064.345 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKKYQYFALALCAALLAG.....BRAGHGHDELAVASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues
Total number of hits satisfying chosen parameters: 283224

Maximum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 73:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2246	95.5	499	B81914	probable periplasm
2	985.5	41.9	474	B83550	serine proteinase
3	862.5	36.7	514	A82581	periplasmic protei
4	735	31.2	523	A97479	probable serine pr
5	735	31.2	523	A12696	serine proteinase
6	716	30.4	493	C97605	probable serine pr
7	716	30.4	514	AE2827	serine proteinase
8	693.5	29.5	513	AD3418	proteinase DO (EC
9	692.5	29.4	513	I40060	serine proteinase
10	683	29.0	497	B81728	serine proteinase,
11	683	29.0	524	AG3328	probable do serine
12	683	29.0	524	AG3328	proteinase do (EC
13	669.5	28.5	488	G81528	serine proteinase,
14	668.5	28.4	488	H8612	DO serine proteina
15	668.5	28.4	488	G72011	DO serine proteina
16	653.5	27.8	457	AG0433	proteinase (EC 3.4
17	649	27.6	481	AH0410	global stress requ
18	648	27.5	466	A64113	heat shock protein
19	639	27.2	474	A13349	proteinase DO (EC
20	638	27.1	511	F97720	periplasmic serine
21	637	27.1	474	I40059	htrA-like protein
22	636.5	27.1	475	AC0528	protease DO precu
23	636	27.0	530	F87590	serine proteinase
24	633.5	26.9	475	S15337	heat shock protein
25	632.5	26.9	455	F82307	proteinase DO VC05
26	631	26.8	429	AD1894	serine proteinase
27	627.5	26.7	455	UC6051	trypsin-like prote
28	623.5	26.5	455	C91142	serine endoprotein
29	617.5	26.2	455	F85987	serine endoprotein

30	617.5	26.2	455	2	AB0909	serine protease (E
31	615.5	26.2	513	2	B71722	probable periplasm
32	614.5	26.1	459	2	F72359	periplasmic serine
33	610.5	25.9	474	2	S45329	proteinase DO (EC
34	610.5	25.9	474	2	E85500	proteinase DO (EC
35	610.5	25.9	474	2	E90649	proteinase DO (EC
36	606.5	25.8	452	2	S77538	serine proteinase
37	604.5	25.7	481	2	D82826	heat shock protein
38	600.5	25.5	416	2	S75445	proteinase htrA (E
39	594.5	25.3	478	2	G84956	proteinase do prec
40	589	25.0	389	2	B83089	AlgW protein PA44
41	582	24.7	476	2	H71936	proteinase DO - He
42	574.5	24.4	443	1	C64647	serine proteinase
43	571	24.3	453	1	B70426	periplasmic serine
44	564	24.0	468	2	A12811	serine proteinase
45	564	24.0	495	2	B97590	htrA protein homol

ALIGNMENTS

RESULT 1									
B81914									
probable periplasmic serine proteinase (EC 3.4.21.-) NMA0710 [imported] - Neisseria men									
C:Species: Neisseria meningitidis									
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001									
C:Accession: B81914									
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Chumcher, C.; Klee, S.R.; Morel									
; Holroyd, S.; Dageis, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,									
Nature 404, 502-506, 2000									
A>Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491.									
A:Reference number: A81775; MUID:20222556; PMID:10761919									
A:Accession: B81914									
A>Status: preliminary									
A:Molecule type: DNA									
A:Residues: 1-499 <PAR>									
A:Cross-references: GB:AL162754; GB:AL157959; NID:97379424; PIDN:CAB8396.1; PID:97379434									
A:Experimental source: serogroup A, strain Z2491									
C:Genetics:									
A:Gene: NMA0710									
C:Superfamily: Helicobacter serine proteinase									
C:Keywords: hydrolase; serine proteinase									
Query Match 95.5%; Score 2246; DB 2; Length 499;									
Best Local Similarity 95.9%; Pred. No. 3,2e-138;									
Matches 444; Conservative 6; Mismatches 13; Indels 0; Gaps 0;									
QY	1	VFKKYQYFALALCAALLAC	ERASFFGADKKERASFEVRIEHTKDDGSVMLLPDFAQL	60					
DB	1	MFKKYQYFALALCAALLAC	ASLACDRAKASFFGADKKERASFEVRIEHTKDDGSVMLLPDFAQL	60					
QY	61	VOSEGPVAVNIQAAPART	ONGSGNAETDSDPLADSPFFYFFFRKLVPMNEIPEEADD	120					
DB	61	VOSEGPVAVNIQAAPART	ONGSGNAETDSDPLADSPFFYFFFRKLVPMNEIPEEADD	120					
QY	121	GGNFGSGFTISKNGVLT	NTHVAVAGSIVKVLNDKREYAKLIGSVQSDVALLKIDA	180					
DB	121	GGNFGSGFTISKNGVLT	NTHVAVAGSIVKVLNDKREYAKLIGSVQSDVALLKIDA	180					
QY	181	TEELPVVKIGNPNKLK	GEWVAALGAPFGDNSTAGIVSAKGSLEPNESYTPFIQTDA	240					
DB	181	TEELPVVKIGNPNKLK	GEWVAALGAPFGDNSTAGIVSAKGSLEPNESYTPFIQTDA	240					
QY	241	INFGNSGGPLFNLKGV	YVINSQIYSSGGFMGISFAPIDVAVNVVAEQLKNTGVQSGQ	300					
DB	241	INFGNSGGPLFNLKGV	YVINSQIYSSGGFMGISFAPIDVAVNVVAEQLKNTGVQSGQ	300					
QY	301	LGVTIOEVSYGLASQF	LDKASGALIKILPGSPAERAGLQAGIVSLDGEIRSSGDL	360					
DB	301	LGVTIOEVSYGLASQF	LDKASGALIKILPGSPAERAGLQAGIVSLDGEIRSSGDL	360					
QY	361	PVWVGALTPEKEVSL	GWKGEITTIKALGNAEHTGASSKTDEAPYTEOQSGTFVES	420					
DB	361	PVWVGALTPEKEVSL	GWKGEITTIKALGNAEHTGASSKTDEAPYTEOQSGTFVES	420					

Db 361 PVMGAIIPGKEVSLGWRKKEEITIKVKLGNAAEHIGASSKTDEAPYTEQSGTSFVS 420
OY 421 AGITLQHTDSSGKHLVVRVSPDAERAGLRHGDEILAVRASP 463
Db 421 AGITLQHTDSSGKHLVVRVSPDAERAGLRHGDEILAVRASP 463

RESULT 2
F83550
serine proteinase MucD precursor PA0766 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_rev1501 15-Sep-2000 #text_change 03-Jun-2002
C:Accession: F83550
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warener, P.; Hickey, M.J.; Bradman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lapidig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathob
A:Reference number: A82950; PMID:20437337; PMID:10984043
C:Accession: F83550
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-474 <STO>
A:Cross-references: GB:AE004511; GB:AE004091; NID:99466646; PIDN:AA04155.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: mucD; PA0766
C:Superfamily: Helicobacter serine proteinase

Query Match 41.9%; Score 985.5; DB 2; Length 474;
Best Local Similarity 52.5%; Pred. No. 1.8e-56;
Matches 220; Conservative 55; Mismatches 111; Indels 33; Gaps 11;

54 LPDFALVQSEBPVAVNI---QAAPARTONGSGMNETDSDPLADSP-PYEFKRLVFN 109
Db LPDFALVQSEBPVAVNI---QAAPARTONGSGMNETDSDPLADSP-PYEFKRLVFN 109
27 LPDFALVQSEBPVAVNI---QAAPARTONGSGMNETDSDPLADSP-PYEFKRLVFN 82

OY 110 MPEIR---QEADDDGGLNFGSGFIISKNGYILTNTHVAVGMSIKVLLNDKREYATKL 166
Db 83 VPRNRGQGRQ---SLGSGFIISNDGIILTNTHVAVGMSIKVLLNDKREYATKL 138
167 SDVQSDVALLKIDATELPELVKIGNPKLPGEWAAIGAPFGDNSTAGIYSAKGRSL 226
Db 139 ADPRSDVALLKIDATELPELVKIGNPKLPGEWAAIGAPFGDNSTAGIYSAKGRSL 197

OY 227 PNESTPPFIQTVDVAINPGNSGGPLFNLKGVVNGINSQIYSSSGFPMGIFSAIPIDVAMN 286
Db 198 PNESTPPFIQTVDVAINPGNSGGPLFNLKGVVNGINSQIYSSSGFPMGIFSAIPIDVAMN 257
287 AEQLKNTGKVGQGLVITIOEVSYGLAOSFGLDKASGALLAKILPGSPARAGLQAGDIV 346
Db 258 ADQLKKAAGVNSGWLGVIOEVKDLAESFGLDKPSALVALVQVDEPAAKGLQVGVYI 317
347 LSLDGEIRSGDLPVVGAIIPGKEVSLGWRKKEEITIKAKLGNAAEHTGASSKTDE 405
OY 318 LSLNGOSINESADLPVLVGMKPGDKINLIVIRNGQKSLMAVGSIPD-----DDEB 370

OY 406 -----APYTEQSGTSFVSAGITLQTH--TDSGKHLVVRVSDA-AERAGLRHGEI 456
Db 371 IASMGAPGERSNRNLGVVADLTAEORSLDIQ--GVVKEVGDGPAVAVGLRPGDVI 428

RESULT 3

A82581
periplasmic proteinase Xp2241 [imported] - Xylella fastidiosa (strain 945c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_rev1501 20-Aug-2000 #text_change 02-Sep-2000
C:Accession: A82581
R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; PMID:20365717; PMID:10910347
A:Note: for a complete list of authors see reference number A59328 below

A:Accession: A82581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <SIM>
A:Cross-references: GB:AE004037; GB:AE003849; NID:99107394; PIDN:AAE8040.1; GSPDB:GN001;
A:Experimental source: strain 945c
R:Stimpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Brienes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carreiro, D.M.; Carreir, H.
as-Neto, E.; Docena, C.; Di-Dorri, H.; Facincant, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohme
J.D.; Unqueira, M.L.; Kemper, E.L.; Kitejina, J.P.; Kiteger, J.E.; Kurumae, E.E.; Laigre
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E.
A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawaaki
A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveira
M.; Tuhakko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:
A:Gene: Xp2241
C:Superfamily: Helicobacter serine proteinase

Query Match 36.7%; Score 862.5; DB 2; Length 514;
Best Local Similarity 46.2%; Pred. No. 1.9e-48;
Matches 194; Conservative 64; Mismatches 129; Indels 33; Gaps 10;

OY 54 LPDFALVQSEBPVAVNIQAAPARTONGSGMNETDSDPLADSP-PYEFKRLVFN 107
Db LPDFALVQSEBPVAVNIQAAPARTONGSGMNETDSDPLADSP-PYEFKRLVFN 106
54 LPDFALVQSEBPVAVNIQAAPARTONGSGMNETDSDPLADSP-PYEFKRLVFN 106

OY 108 PNMPEIRPEADDDG---NFGSGFIISKNGYILTNTHVAVGMSIKVLLNDKREYATKL 164
Db 107 PNMPEIRPEADDDG---NFGSGFIISKNGYILTNTHVAVGMSIKVLLNDKREYATKL 164
107 PNMPEIRPEADDDG---NFGSGFIISKNGYILTNTHVAVGMSIKVLLNDKREYATKL 164

OY 165 IGSVDQSDVALLKIDATELPELVKIGNPKLPGEWAAIGAPFGDNSTAGIYSAKGR 224
Db 165 IGSVDQSDVALLKIDATELPELVKIGNPKLPGEWAAIGAPFGDNSTAGIYSAKGR 223
225 SLPNES-YTPFIQTVDVAINPGNSGGPLFNLKGVVNGINSQIYSSSGFPMGIFSAIPIDV 283
OY 224 STSDDRVYPIQTVDVAINPGNSGGPLFNLKGVVNGINSQIYSSSGFPMGIFSAIPIDV 283
284 MNVAEQLKNTGKVGQGLVITIOEVSYGLAOSFGLDKASGALLAKILPGSPARAGLQAG 343
OY 284 MNVAEQLKNTGKVGQGLVITIOEVSYGLAOSFGLDKASGALLAKILPGSPARAGLQAG 343
284 MNVAEQLKNTGKVGQGLVITIOEVSYGLAOSFGLDKASGALLAKILPGSPARAGLQAG 343

OY 344 DIVLSLDGEIRSGDLPVVGAIIPGKEVSLGWRKKEEITIKAKLGNAAEHTGASS 401
Db 344 DIVLSLDGEIRSGDLPVVGAIIPGKEVSLGWRKKEEITIKAKLGNAAEHTGASS 402
344 DIVLSLDGEIRSGDLPVVGAIIPGKEVSLGWRKKEEITIKAKLGNAAEHTGASS 402
OY 402 KTDAPYTEQSGTSFVSAGITLQHTDSSGKHLVVRVSDAERAGLRHGEILAVRA 461
Db 403 KTDAPYTEQSGTSFVSAGITLQHTDSSGKHLVVRVSDAERAGLRHGEILAVRA 452

RESULT 4

A97479
probable serine proteinase homolog precursor [imported] - Agrobacterium tumefaciens (str
C:Species: Agrobacterium tumefaciens
C:Date: 30-Sep-2001 #sequence_rev1501 30-Sep-2001 #text_change 17-May-2002
C:Accession: A97479
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Woliam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum

A:Reference number: A97359; PMID:11743194
A:Accession: A97479
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-523 <KUR>
A:Cross-references: GB:AE007869; PIDN:AAK86786.1; PID:G1515598; GSPDB:GN00169

C:Genetics:
 A:Gene: AGR_C_1792
 A:Map position: circular chromosome
 C:Superfamily: Helicobacter serine proteinase

Query Match 31.2%; Score 735; DB 2; Length 523;
 Best Local Similarity 37.0%; Pred. No. 3.7e-40;

Matches 168; Conservative 82; Mismatches 154; Indels 50; Gaps 9;

QY 36 SFVERIHTKDDGVSMLPDPFAQLVQSEGPVAVNIQAAPAPRTONGSNAETDS-----90

DB 44 SFAAPVETAPQ-----VPSFANVVDAVPAVSV-----RVQSNVQPASDSSNFSF 91

QY 91 -----DPLADSDPFYEFKRLVPMNPBEPQSEADGGLN-----FGSGFI 130

DB 92 NFGGRGDDQLPDDHPLKRFKFEFGQ-----NQDRSDGPNRHRDQKGPLRPVAGSGGF 146

QY 131 ISKNGYILTNTHVAVAGMSIKVLNDKREYTKLIGSDVQSVALLKIDATEELPVYKIG 190

DB 147 ISEGGYVVTNNHVVDDGSAITVVMNDGTELEKALVGRDPRDALLKVDVNRKFTYVYKFA 206

QY 191 NPKNLKGEVVAALGAPFGFNSVTAGIVSAKGRSLPNESYTPFIQTDAINPNSGGPL 250

DB 207 DDTKIRVGDVAVAGNPGFGLGTVTSGIISARGRIDSGPYDYLQIDAAVNRKNSGGA 266

QY 251 FNKGVGVGINSQIYSRSGFMGISFAIPIDVANNVABQLKNTGKVGQGLGVIIQEVSY 310

DB 267 FNNGEVVGINATLFPSPGGNVGIAFAIPSSVAKVDIADLQKDGKVERGWLGVQIQPVSK 326

QY 311 GLAQSGLDKASGALITAKILPGSPAERAGLQAGDIVLSLDGEIRSSGDLPMVGAITPG 370

DB 327 DIAESLGIAEAKGALVVSPOSGSPDKAGIKQGDITLVANGDPVARDLSRRIGMAPN 386

QY 371 KEVSLGVWRKGEITITKAKLGN--AAEHTGASSKTDEAPYTEQSGTFVSAGITLQTH 428

DB 387 SKYEISLMRGKSGSVTVTLGDLTSDASAKATPSONDKGSGSSSEKVLSTGLTVSPS 446

QY 429 TDSGKHLVVVRV---SDAERAGLRHGDILLAV 459

DB 447 DDGNG--LAITVDVDPDSDAAR-GLTKGEKITSV 477

RESULT 5
 A12696
 serine proteinase DO-like proteinase dop [imported] - Agrobacterium tumefaciens (strain

C:Species: Agrobacterium tumefaciens
 C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 17-May-2002

C:Accession: A12696
 R:Wood, D.W.; Secubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, I.

RP, F.; Kometo, P.; Grant, C.; Guentherer, D.; Kutayavin, T.; Levy, R.; Li, M.; Mccllell

Sequence 294, 2317-2323, 2001
 A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krepan, W.; Perry, M.; Gordon-Kamm,

ster, B.W.
 A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.

A:Reference number: AB2577; PMID:11743193
 A:Accession: A12696

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-523 <KUR>
 A:Cross-references: GB:AE008688; PIDN:AL41991.1; PID:917739363; GSPDB:GN00186

A:Experimental source: strain C58 (Dupont)
 C:Genetics:

A:Gene: dop
 A:Map position: circular chromosome

C:Superfamily: Helicobacter serine proteinase

Query Match 31.2%; Score 735; DB 2; Length 523;
 Best Local Similarity 37.0%; Pred. No. 3.7e-40;

Matches 168; Conservative 82; Mismatches 154; Indels 50; Gaps 9;

QY 36 SFVERIHTKDDGVSMLPDPFAQLVQSEGPVAVNIQAAPAPRTONGSNAETDS-----90

DB 44 SFAAPVETAPQ-----VPSFANVVDAVPAVSV-----RVQSNVQPASDSSNFSF 91

QY 91 -----DPLADSDPFYEFKRLVPMNPBEPQSEADGGLN-----FGSGFI 130

DB 92 NFGGRGDDQLPDDHPLKRFKFEFGQ-----NQDRSDGPNRHRDQKGPLRPVAGSGGF 146

QY 131 ISKNGYILTNTHVAVAGMSIKVLNDKREYTKLIGSDVQSVALLKIDATEELPVYKIG 190

DB 147 ISEGGYVVTNNHVVDDGSAITVVMNDGTELEKALVGRDPRDALLKVDVNRKFTYVYKFA 206

QY 191 NPKNLKGEVVAALGAPFGFNSVTAGIVSAKGRSLPNESYTPFIQTDAINPNSGGPL 250

DB 207 DDTKIRVGDVAVAGNPGFGLGTVTSGIISARGRIDSGPYDYLQIDAAVNRKNSGGA 266

QY 251 FNKGVGVGINSQIYSRSGFMGISFAIPIDVANNVABQLKNTGKVGQGLGVIIQEVSY 310

DB 267 FNNGEVVGINATLFPSPGGNVGIAFAIPSSVAKVDIADLQKDGKVERGWLGVQIQPVSK 326

QY 311 GLAQSGLDKASGALITAKILPGSPAERAGLQAGDIVLSLDGEIRSSGDLPMVGAITPG 370

DB 327 DIAESLGIAEAKGALVVSPOSGSPDKAGIKQGDITLVANGDPVARDLSRRIGMAPN 386

QY 371 KEVSLGVWRKGEITITKAKLGN--AAEHTGASSKTDEAPYTEQSGTFVSAGITLQTH 428

DB 387 SKYEISLMRGKSGSVTVTLGDLTSDASAKATPSONDKGSGSSSEKVLSTGLTVSPS 446

QY 429 TDSGKHLVVVRV---SDAERAGLRHGDILLAV 459

DB 447 DDGNG--LAITVDVDPDSDAAR-GLTKGEKITSV 477

RESULT 6
 C97605
 probable serine proteinase DO-like precursor [imported] - Agrobacterium tumefaciens (stre

C:Species: Agrobacterium tumefaciens
 C:Date: 30-Sep-2001 #sequence_revision 30-Sep-2001 #text_change 11-Jan-2002

C:Accession: C97605
 R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,

A.; Liu, F.; Wollam, C.; Allinger, M.; Doughy, D.; Scott, C.; Lappas, C.; Markelz, B.;

Science 294, 2323-2328, 2001
 A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tume

A:Reference number: A97359; PMID:11743194
 A:Accession: C97605

A:Status: preliminary
 A:Molecule type: DNA

A:Residues: 1-493 <KUR>
 A:Cross-references: GB:AE007869; PIDN:AAK87796.1; PID:915157169; GSPDB:GN00169

C:Genetics:
 A:Gene: AGR_C_3700

A:Map position: circular chromosome

Query Match 30.4%; Score 716; DB 2; Length 493;
 Best Local Similarity 36.4%; Pred. No. 5.8e-39;

Matches 170; Conservative 75; Mismatches 146; Indels 76; Gaps 10;

QY 48 GSVSMILPDPFAQLVQSEGP-----AVNIOAAPAPRTONGSNAETDSPLAD 95

DB 3 GSLS---APVABARQSHGPEVSADLAERPLDVAVINISTQNKTE-CKEPV---PKLPE 55

QY 96 SDPFYEFKRLVPMNPBEPQSEADGG--LNFSGFTIISKNGYILTNTHVAVAGMSIKY 152

DB 56 GSPFOEFK-----DYFDSQKPEGGEKVNLSGSGFVIDPAGYVVTNNHNVIGADAIEV 108

QY 153 LLNDKREYTKLIGSDVQSVALLKIDATEELPVYKIGPKNLKGEVVAALGAPFGFN 212

DB 109 IFPNQSKLAKITLVGDTITDLSVLKVEKPTPLKAYKFGDSRNRIGDWMVAVGNPFLGLG 168

QY 213 SVTAVIVSAKGRSLPNESYTPFIQTDAINPNSGGPLFNKGVGVGINSQIYSRSGFM 272

DB 169 SLTVGVISARGNINAVAGPYDNLQTDALINKNSGGPLFNMGVEVIGINTAIISPSGGS 228

QY 273 GISFAIPIDVANNVABQLKNTGKVGQGLGVIIQEVSYGLAQSGLDKASGALITAKILPG 332

Db 229 GIGFAVPTLQNIIVQOLIEFGETRGWLGVRVQPTDDVAASIGMSAKALISGVAKG 288

Qy 333 SPAERAGLQAGDIVLSLDGGEIRSSGDLPMVVGAIITPGKEVSLGWRKGEIITIKAKLGN 392

Db 289 GFVENGPICQAGDVVLFKDGKIDINEMRDLRIIVASPVGKEVDVVYVDGKEETVKVLLGQ 348

Qy 393 AAEHTGASSKTDEAPYTEQSG-----TFVSESAGITLQTHTDSGK 434

Db 349 LQDTTDEKASTDD-PCGEBDGGWVAAPDDKGDGDAQDQTPPEVREAPQTV-----LGM 401

Qy 435 HLTV-----VRSDAERAGLRHGDDELAV 459

Db 402 NLVVLNSNELRTKGIASVEGVLVASVDPSPAEQKMKAGDIVVEV 448

RESULT 7
AE2827
serine proteinase DO-like precursor htra (imported) - Agrobacterium tumefaciens (strain
Accession: AE2827
Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 11-Jan-2002
C:Accession: AE2827
R:Wood, D.W.; Saebul, J.C.; Kaul, R.; Monke, D.; Chen, L.; Wood, G.E.; Chen, Y.; Moo, I.
erage, G.; Gille, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCell
: Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
A:er, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: AB5577, PMID:11743193
A:Accession: AE2827
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-514 <KUR>
A:Cross-references: GB:AE008688; PIDN:AA143035.1; PID:G17740500; GSPDB:GN00186
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: htra
A:Map position: circular chromosome

Query Match 30.4%; Score 716; DB 2; Length 514;
Best Local Similarity 36.4%; Pred. No. 6,1e-39;
Matches 170; Conservative 75; Mismatches 146; Indels 76; Gaps 10;

Qy 48 GSVMLPDPFAQLVQSEGP-----AVVNIQAAPRTONGSGNAETDSDPLAD 95

Db 24 GSLS---APVAAQAQHGPESSVADLAEPFLDVAVNISTSQNVKTE-CKGPV---PKLPE 76

Qy 96 SDPFFEFFKRLVPMMEIIPQEEADDCG--LNFSGGFIISKNGIYILNTYVAVAGMSIKV 152

Db 77 GSPFQEFFK-----DYFDSQKPEEGEKVNSLGSQFVIDPAGVYVNNHVIIEGADALEV 129

Qy 153 LINDKREYAKLIGSVOSDVALIKTIDATELPLVVKIGNPKLKPGEVVAIGAIFGDN 212

Db 130 ITPNRSKLTATLVGTITKTDLSTLVKBPKTPLKAVKFGDSRSMRIGWMAVAGNPFELGG 189

Qy 213 SVTAGIVSAGKSLPNESYTPPIQTDAIINPNSGGFLFNLKQGVVINSQIVSRSGFM 272

Db 190 SLTVGIVSARGRNINAGPYDNFIQTDAIINKNSGGFLFNMKGEVIGINIAIISPSSSI 249

Qy 273 GISFAIPVAMVVAQLKTKGVQSGQLVITIQEVSYGLAQSGFLDKASGALIAKLIPG 332

Db 250 GIGFAVPTLQNIIVQOLIEFGETRGWLGVRVQPTDDVAASIGMSAKALISGVAKG 309

Qy 333 SPAERAGLQAGDIVLSLDGGEIRSSGDLPMVVGAIITPGKEVSLGWRKGEIITIKAKLGN 392

Db 310 GEVENGPICQAGDVVLFKDGKIDINEMRDLRIIVASPVGKEVDVVYVDGKEETVKVLLGQ 369

Qy 393 AAEHTGASSKTDEAPYTEQSG-----TFVSESAGITLQTHTDSGK 434

Db 370 LQDTTDEKASTDD-PCGEBDGGWVAAPDDKGDGDAQDQTPPEVREAPQTV-----LGM 422

Qy 435 HLTV-----VRSDAERAGLRHGDDELAV 459

Db 423 NLVVLNSNELRTKGIASVEGVLVASVDPSPAEQKMKAGDIVVEV 469

RESULT 8
AD3418
proteinase DO (EC 3.4.21.-) (imported) - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
A:Accession: AD3418
R:DeVechio, V.G.; Kapural, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova, I.
: Mazur, M.; Goldeman, E.; Selkov, E.; Elizer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AD3418
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-513 <KUR>
A:Cross-references: GB:AE00917; PIDN:AA152511.1; PID:G17983322; GSPDB:GN00190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BME11330
A:Map position: I
A:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 29.5%; Score 693.5; DB 2; Length 513;
Best Local Similarity 37.4%; Pred. No. 1,8e-37;
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

Qy 9 ALAALCALAGCEKAGSFFGADKKEASFERIEHTKDGVSMLLPD--FAQLVQSEG 65

Db 14 AAVALSAL-----AGAFVVTGFLGALNEMRAE-----AVHTVPPQAGADIVLEKVR 61

Qy 66 PAVVNIQAAPRTONGSGN--AETDSDPLADSPFFEFF-----KRLV 107

Db 62 PAVSVRVKQDVQETSNGPQFPFGPDQLPDGHLKRFPRDFGMEPRGDSRSDNRK 121

Qy 108 PMMEIIPQEEADDCGLNFGSGFIISKNGIYILNTYVAVAGMSIKVILINDKREYAKLIGS 167

Db 122 ANKRPPGHER-----VAQSGFVISEDDYVNTNNVSDGDAIYTVLDGTELDKLTIA 177

Qy 168 DVQSDVALIKTIDATE-ELPVYKIGNPKLKPGEVVAIGAIFGPNSTAGIVSAGKSL 226

Db 178 DPTDVLAKINAKRKRVVYAFGDNDKRVGDWVAAGNPFGLGTVTSGIVSARGDI 237

Qy 227 PNESTYPIQTDVAIINPNSGGFLFNLKQGVVINSQIVSRSGFMISFAIPIDVAMV 286

Db 238 GAGPYDDPTQIDAIVNKNSSGSPAPDLSGEVIGINTALFSPSGSVGIAFALPSTAKOV 297

Qy 287 AEQLKNTGVQSGQLVITIQEVSYGLAQSGFLDKASGALIAKLIPGSPAERAGLQAGDIV 346

Db 298 VDQILKKGVSRRGVIQVQIQTDAIASLGAELKAIIVASPODDGPAKAGIRAGDIV 357

Qy 347 LSLDGEIRSSGDLPMVVGAIITPGKEVSLGWR--KGEIITK-AKLGNAAEHTGASSKT 403

Db 358 TAVNGETVQDPRDLARKVANIAPGEKALTVMRKKAKEINVTIAMPDKKSGSQSND 417

Qy 404 DEAPYTEQSGTFVSESAGITLQTHTDSGKHLVYVR---SDAERAGLRHGDDELAV 459

Db 418 NDGGQGE-----TLDSGLTVVPSD--GKGVVTVDDPDSADAR-GIRSGDIVSV 467

RESULT 9
I40060
serine proteinase (EC 3.4.21.-) htra, temperature-inducible [validated] - Brucella abortu
C:Species: Brucella abortus
C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 26-May-2000
A:Accession: I40060
R:Tatum, F.M.; Cheville, N.F.; Morfitt, D.
Microb. Pathog. 17, 23-36, 1994
A:Title: Cloning, characterization and construction of htra and htra-like mutants of Bruc
A:Reference number: I40059; MUID:95165990; PMID:7861951

A:Accession: 140060
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-513 <RES>
A:Cross-references: EMBL:U07352; NID:g497156; PIDN:AAA70164.1; PID:g497157
C:Genetics:
A:Gene: htra
C:Function:
A:Description: EC 3.4.21.-; serine proteinase [validated, MUID:95165990]; permits growth
C:Superfamily: Helicobacter serine proteinase
C:Keywords: hydrolase; serine proteinase

Query Match 29.4%; Score 692.5; DB 2; Length 513;
Best Local Similarity 37.4%; Pred. No. 2e-37;
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

9 ALAALCAALLAGCEKAGSPFGADKKEASFVERIEHTKDDGSYMLP---DFAQLVQSEG 65
14 AAVALLNAL-----AGAFVYTGPGALNEARAE-----AAHVTPPQAGFDLVKVR 61
66 PAVVNIQAAPAPRTONGSGN---AETSDPLADSDPFYEF-----KRLV 107
62 PAVVSVRVKKDVOETSNRGPFGPFGPDQLPDGHPDKFFPDPGMEPRGDSNDNRGK 121
108 PMNPETPOEADDCGLAFSGSFITSKNGYILTNTYVAGKSIKVLANDKREYATLIGS 167
122 ANKPRGHERP---VAQSGSFVISEDGVYVNNHVSNGDAYTVVLLDGTETLDAKLGA 177
168 DVOSDVALKIIDATE-ELAPVVKIGNPKMLKPEWYAAIGAPEGFDSVYAGIVSANGRS 226
178 DPTIDLAVALKINPKRKFVYVYAFPGDNDKVRGDMVAANPGLGTVISGIVSANGRD 237
227 PNESTYPTFIOTDVAINPNSGGLPFLNKGQVNVINSQIYSRSGFNGISFAIPDIVAMN 286
238 GAGPYDDFIDIDAIVNNGNSGPAFDLSGEVIGINTAIFSPGSGVIGIAFATISSRAKV 297
287 AEOLKNTGKYQVQGLGYITQEVSYGLAOSFGIDKASGALIAKILPGSPAPRAGLQAGDI 346
298 VQLIKKGSVERGWIQVQIPVTKDIAASLGLAEKGAIVASPDQDGPAPAKAIGKAGDI 357
347 LSLDGEIRSSGDLPWVGAITTPGKEVSLGWVR--KGEIITIK-ALGNAAERTGASKT 403
358 TANGETVQPRDLARKYANLAPSEKALTWKRKKAEIINVTIAMPNDRKSGSQSND 417
404 DEAPYTEQSGCTFSVESAGITLTQHTDSSGKHLVVVR---SDAARAGLRHGDEITAV 459
418 NDDGQSE-----TLDSYGLTVPSSE--GKGVVTVDDPDSDADNR-GIRSDVIVSV 467

LT 10
28

serine proteinase, Htra/DegQ/DegS family TC0210 [imported] - Chlamydia muridarum (strain
C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 21-Jul-2000
C:Accession: B81728
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gilm, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: AB1500; MUID:20150255; PMID:10684935
A:Accession: B81728
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-497 <RES>
A:Cross-references: GB:AB002286; GB:AB002160; NID:g7190247; PIDN:AAF39082.1; PID:g719022
C:Genetics:
A:Experimental source: strain N19g (MoPn)
C:Gene: TC0210
C:Superfamily: Helicobacter serine proteinase

Query Match 29.0%; Score 683; DB 2; Length 497;
Best Local Similarity 37.4%; Pred. No. 8.1e-37;
Matches 173; Conservative 77; Mismatches 154; Indels 58; Gaps 11;

QY 28 FGADKKKSAFVERIHTKDDQSVSM--LLPD----FAOLVQSSGPAVNNIOAAPARTQ- 80
 Db 22 YSAPKKDSTGTCLASQSDRELQSDBLKEVSRGFSKRAAQAATPGVYIENFPKTSQA 81
 QY 81 -----NSGNAETDSDPLADSDPFYEFFPKRL--VPNMPEIQEADQGLNFGSGFIISK 133
 Db 82 IASPNKKRGFOENPRDFYND-----EFFNRFGPLRPHREGPRRQQRDAVR--GTGFIYSE 134
 QY 134 NGYIITNTTHVVAQWGSIKYLLNDKREYTKALIGSDVQSDVALIKIDATEELPVVKIGNPK 193
 Db 135 DGYVVTNNHHVVEDACKIHVTLHDGQKYTKAIIGLDPKTDLAVIKIQK-KMLPFLTFNSD 193
 QY 194 NLKPEWVAALGAPRGFNDSTAGIVSAKGRS-LPNESTYPTQTVAINPGNSGGFLFN 252
 Db 194 QLOIGDMSTIAGNPPGLOATVTVGVISAKGRQDLHIVDEDFIQTDALINPGNSGGFLFN 253
 QY 253 LKGQVVGINSQIYSRSGGFMGISFAIPIDVANNVAEQLKNTGKVRQGLGVIIIOEVSYGL 312
 Db 254 IDGQYIGVNTAIVSGSGGYIGIGFAPISLMARKVIDQLISDQVTRGFLGVTTLQPIDSEL 313
 QY 313 AQSFLDRAAGALLIKILPGSPAERAGLOAGDIVLSLDGEIRSSGDLPMVGAITTEKE 372
 Db 314 AACYKLEKYGALLIDVVKSGSPAERAGLQEDVITVAYNGKEVESLSALRNAISLMPGTR 373
 QY 373 VSLGVRKGEETITRAKKGNMAEHTGASKTDPAAPTEQO---SGFVSYESAGITLQTH 428
 Db 374 VLKVRKRGKFT-----EIPATVQIIPAEQVSALQKMGVRVQNL 413
 QY 429 T-----DSSGKHLVVVRVSDPAERAGLRHGDILLAV 459
 Db 414 TPETCKKLGASDTRGIFVSVSEAGSPASAGVPPQLILAV 455

RESULT 11
 H71465
 probable do serine proteinase - Chlamydia trachomatis (serotype D, strain UW3/Cx)
 C:Species: Chlamydia trachomatis
 C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #ext_change 08-Oct-1999
 C:Accession: H71465
 R:Stephens, R.S.; Kallan, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,
 Science 282, 754-759, 1998
 A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis
 A:Reference number: A71570; MUID:99000809; PMID:9784136
 A:Accession: H71465
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-497 <ARN>
 A:Cross-references: GB:AE001355; GB:AE001273; NID:G3329292; PIDD:ACC68420.1; PID:G3329292;
 A:Experimental source: serotype D, strain UW-3/Cx
 C:Genetics:
 A:Gene: htrA
 C:Superfamily: Helicobacter serine proteinase

Query Match 29.0%; Score 683; DB 2; Length 497;
 Best Local Similarity 37.0%; Pred No. 8, 1e-37;
 Matches 171; Conservative 75; Mismatches 158; Indels 58; Gaps 11;

QY 28 FGADKKKSAFVERIHTKDDQSVSM--LLPD----FAOLVQSSGPAVNNIOAAPARTQ- 80
 Db 22 YSAPKKDSTGTCLASQSDRELQSDBLKEVSRGFSRVAAKATPGVYIENFPKTNQA 81
 QY 81 -----NSGNAETDSDPLADSDPFYEFFPKRL--VPNMPEIQEADQGLNFGSGFIISK 133
 Db 82 IASPNKKRGFOENPRDFYND-----EFFNRFGPLRPHREGPRRQQRDAVR--GTGFIYSE 134
 QY 134 NGYIITNTTHVVAQWGSIKYLLNDKREYTKALIGSDVQSDVALIKIDATEELPVVKIGNPK 193
 Db 135 DGYVVTNNHHVVEDACKIHVTLHDGQKYTKAIIGLDPKTDLAVIKIQK-ETLPLTFNSD 193
 QY 194 NLKPEWVAALGAPRGFNDSTAGIVSAKGRS-LPNESTYPTQTVAINPGNSGGFLFN 252
 Db 194 QLOIGDMSTIAGNPPGLOATVTVGVISAKGRQDLHIVDEDFIQTDALINPGNSGGFLFN 253
 QY 253 LKGQVVGINSQIYSRSGGFMGISFAIPIDVANNVAEQLKNTGKVRQGLGVIIIOEVSYGL 312
 Db 254 IDGQYIGVNTAIVSGSGGYIGIGFAPISLMARKVIDQLISDQVTRGFLGVTTLQPIDSEL 313
 QY 313 AQSFLDRAAGALLIKILPGSPAERAGLOAGDIVLSLDGEIRSSGDLPMVGAITTEKE 372
 Db 314 AACYKLEKYGALLIDVVKSGSPAERAGLQEDVITVAYNGKEVESLSALRNAISLMPGTR 373
 QY 373 VSLGVRKGEETITRAKKGNMAEHTGASKTDPAAPTEQO---SGFVSYESAGITLQTH 428
 Db 374 VLKVRKRGKFT-----EIPATVQIIPAEQVSALQKMGVRVQNL 413
 QY 429 T-----DSSGKHLVVVRVSDPAERAGLRHGDILLAV 459
 Db 414 TPETCKKLGASDTRGIFVSVSEAGSPASAGVPPQLILAV 455

A:Experimental source: strain J138

C:Genetics:

A:Gene: htra

C:Superfamily: Helicobacter serine proteinase

Query Match 28.4%; Score 668.5; DB 2; Length 488;

Best Local Similarity 36.1%; Pred. No. 6.9e-36;

Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;

QY 10 LAALCAALLAGCEKAGSFFGADKKEASFYERIEHTKDDGSVSMILPDPFQVQSEGPAVY 69

Db 10 LAVLVGSSLLALPLPSGQAVG--KKEs---RVSELPODVLKKEISGSGFSKVAIKAPAVY 63

QY 70 NIQAAPAR--TQNSGNAETSDPLADDPY-EFFKRL--VPMNPEIPQ-EEADGGGL 123

Db 64 YIESFPKSAQVTHPSRGRGYPENPF--DYFNDEFNRFPGLPQREKQSKAVR--- 117

QY 124 NFGSGFIISKNGYILTNTHVAVGMSIKVLLNDKREYTKLIGSDVQSDVALLKIDATEE 183

Db 118 --GTGFLVSPDGYIVTNHNVVEDTGKIHVTLHDGQKYPATYIGLDPKTDIAVIKI-KSQN 174

QY 184 LPVVKIGNPKLKEGEVVAIIGAPFGFDSVTAGIVSAKRS-LPNESYTPPIQTDVAIN 242

Db 175 LPYLSFGNSDHLKVGDMAIAGNPFGLQATVTGVIISAKGRNQLHIADEFDIQTDAAIN 234

QY 243 PGNSSGFLFNKGVVGINSOIYRSRSGFMGISPAIPIDVAMNVAEQKNTGKVQRGQUG 302

Db 235 PGNSSGFLNIDGOVIGVNTAIVSGSGYIGIIFSLPMNRITDQLIRDQVTRGFLG 294

QY 303 VTIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLOAGDIVLSLDGEIRSSGDLVY 362

Db 295 VTLQPIDAEIAACVLEKVGALVTDVKGSPADKAGLKQEDVITIAVNGKEVDSLMPRN 354

QY 363 MYGAIITPGEKVSIGVWRKGEBITTKAKLGNAAEHTGASSKTDEAPYTEOQSGTFPSVESAG 422

Db 355 AVSLMNPEDTRIVLKVREKGVIEIPVTVSQAPKEDGMS-----ALQRVG 398

QY 423 ITLQTHDSSGKHU-----VVVRVSDAERAGLRHGDILLAV 459

Db 399 IRVQNLTPETAKKLGIAPEYTKGILLIISVEPGSVAASSGIAPQULILAV 446

RESULT 15

G72011

do serine proteinase - Chlamydia pneumoniae (strain CWL029)

C/Species: Chlamydia pneumoniae, Chlamydia pneumoniae

C/Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 05-May-2000

C/Accession: G72011

R. Kallman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;

J. Infect. Dis. 1999; 179: 1385-1389, 1999

Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.

A/Reference number: A72000; MUID:99206606; PMID:10192388

A/Accession: G72011

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-488 <ARN>

A/Cross-references: GB:AE001678; GB:AE001363; NID:94377301; PIDN:AAD19116.1; PID:9437730

A/Experimental source: strain CWL029

C/Genetics:

A:Gene: htra

C:Superfamily: Helicobacter serine proteinase

Query Match 28.4%; Score 668.5; DB 2; Length 488;

Best Local Similarity 36.1%; Pred. No. 6.9e-36;

Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;

QY 10 LAALCAALLAGCEKAGSFFGADKKEASFYERIEHTKDDGSVSMILPDPFQVQSEGPAVY 69

Db 10 LAVLVGSSLLALPLPSGQAVG--KKEs---RVSELPODVLKKEISGSGFSKVAIKAPAVY 63

QY 70 NIQAAPAR--TQNSGNAETSDPLADDPY-EFFKRL--VPMNPEIPQ-EEADGGGL 123

Db 64 YIESFPKSAQVTHPSRGRGYPENPF--DYFNDEFNRFPGLPQREKQSKAVR--- 117

QY 124 NFGSGFIISKNGYILTNTHVAVGMSIKVLLNDKREYTKLIGSDVQSDVALLKIDATEE 183

Db 118 --GTGFLVSPDGYIVTNHNVVEDTGKIHVTLHDGQKYPATYIGLDPKTDIAVIKI-KSQN 174

QY 184 LPVVKIGNPKLKEGEVVAIIGAPFGFDSVTAGIVSAKRS-LPNESYTPPIQTDVAIN 242

Db 175 LPYLSFGNSDHLKVGDMAIAGNPFGLQATVTGVIISAKGRNQLHIADEFDIQTDAAIN 234

QY 243 PGNSSGFLFNKGVVGINSOIYRSRSGFMGISPAIPIDVAMNVAEQKNTGKVQRGQUG 302

Db 235 PGNSSGFLNIDGOVIGVNTAIVSGSGYIGIIFSLPMNRITDQLIRDQVTRGFLG 294

QY 303 VTIQEVSYGLAQSFGLDKASGALIAKILPGSPAERAGLOAGDIVLSLDGEIRSSGDLVY 362

Db 295 VTLQPIDAEIAACVLEKVGALVTDVKGSPADKAGLKQEDVITIAVNGKEVDSLMPRN 354

QY 363 MYGAIITPGEKVSIGVWRKGEBITTKAKLGNAAEHTGASSKTDEAPYTEOQSGTFPSVESAG 422

Db 355 AVSLMNPEDTRIVLKVREKGVIEIPVTVSQAPKEDGMS-----ALQRVG 398

QY 423 ITLQTHDSSGKHU-----VVVRVSDAERAGLRHGDILLAV 459

Db 399 IRVQNLTPETAKKLGIAPEYTKGILLIISVEPGSVAASSGIAPQULILAV 446

Search completed: July 11, 2003, 10:33:59
Job time : 44 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:15:32 ; Search time 12 Seconds
(without alignments)
1607.208 Million cell updates/sec

Title: US-09-388-090-4
2353

Perfect score: 1 VFKYQYFALALCALIAG.....ERAGLRHDEILAVRASPRQ 465

Sequence: 1 VFKYQYFALALCALIAG.....ERAGLRHDEILAVRASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 112892 seqs, 41476328 residues

112892

Total number of hits satisfying chosen parameters:

112892

Minimum DB seq length: 0

112892

Post-processing: Minimum Match 0%

112892

Maximum Match 100%

112892

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query	Match Length	ID	Description
1	747	31.7	504	1	DEGP RHIME
2	731.5	31.1	503	1	DEGP BARHE
3	692.5	29.4	513	1	DEGP BRUB
4	683	29.0	497	1	DEGP CHLMU
5	683	29.0	497	1	DEGP CHLTR
6	668.5	28.4	488	1	DEGP CHLTP
7	648	27.5	466	1	HMOA HAHIN
8	638	27.1	508	1	DEGP RICCN
9	633.5	26.9	475	1	DEGP SALTY
10	627.5	26.7	455	1	DEGP ECOLI
11	615.5	26.2	513	1	DEGP RICPR
12	610.5	25.9	474	1	DEGP ECOLI
13	594.5	25.3	478	1	DEGP BUCAL
14	571.5	24.3	478	1	DEGP BUCAP
15	502.5	21.4	449	1	HRAA_BACU
16	470	20.0	437	1	DEGI_ARATH
17	459.5	19.5	458	1	YVTA_BACU
18	448	19.0	480	1	HRAI_HUMAN
19	447.5	19.0	355	1	DEGS_HUMAN
20	436.5	18.6	413	1	HRAA_LACHS
21	435.5	18.5	480	1	HRAI_MOUSE
22	435	18.5	448	1	DEGS_HUMAN
23	428	18.2	340	1	DEGS_HUMAN
24	414.5	17.6	453	1	HRAA_HUMAN
25	414	17.6	453	1	HRAA_HUMAN
26	412.5	17.5	476	1	YVTA_BACU
27	412	17.5	400	1	HRAA_MOUSE
28	411.3	17.5	458	1	HRAA_MOUSE
29	388	16.5	460	1	HRAA_MOUSE
30	353	15.0	460	1	HRAA_MOUSE
31	344.5	14.6	321	1	HMOA_ARATH
32	230	9.8	630	1	YVTA_MOUSE
33	198.5	8.4	997	1	YVTA_MOUSE

34	137.5	5.8	767	1	DLG4_HUMAN	P78352 homo sapien
35	131	5.6	724	1	DLG4_RAT	P31016 rattus norv
36	130	5.5	724	1	DLG4_MOUSE	P62108 mus musculu
37	128	5.4	2021	1	OMPA_RICCN	Q52657 rickettsia
38	125	5.3	430	1	YKJ5_STRCO	Q9K980 streptomyce
39	118	5.0	280	1	ETA_STANU	P09331 staphylococ
40	115.5	4.9	837	1	APBI_HUMAN	Q02410 homo sapien
41	115	4.9	335	1	YG93_CLOPE	Q8XJ12 clostridium
42	115	4.9	568	1	YEIL_DROMD	Q9GP71 drosophila
43	115	4.9	568	1	YEIL_DROSU	Q02437 drosophila
44	114.5	4.9	911	1	DLG1_RAT	Q62696 rattus norv
45	114	4.8	904	1	DLG1_HUMAN	Q12959 homo sapien

ALIGNMENTS

RESULT 1
DEGP_RHIME STANDARD; PRT; 504 AA.
AC Q52894;
AT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-).
GN DEGP1 OR DEGP OR R01021 OR SMC02365.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=96146524; PubMed=8550509;
RA Glazebrook J., Ichige A., Walker G.C.;
RT Genetic analysis of Rhizobium meliloti bacA-phoA fusion results in
RT identification of degP, two loci required for symbiosis are closely
RT linked to degP.";
RL J. Bacteriol. 178:745-752(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boistard P., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Goffre T., Goffeau A., Kahn D., Kiss E., Lelaire V., Maury D.,
RA Pohl T., Portetelle D., Puhler A., Purnelle B., Rampersingh U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>
CC or send an email to license@ebi.ac.uk).
DR EMBL: U31512; AAC3669.1; ALT_INIT.
DR EMBL: AF591785; CAC45593.1; -.
DR MROPS; S01.273; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR000126; Ser protease V8.
DR InterPro; IPR001254; Ser protease V8.
DR Pfam; PF000089; trypsin_1.
DR Pfam; PF00595; PDZ_2.
DR PRINTS; PR00834; PROTEASES2C.

DR PRINTS, PRO0839, V8PROTEASE.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 KW Hydrolyase; Serine protease; Periplasmic; Repeat; Signal;
 KW Complete proteome.
 FT SIGNAL 1 26 POTENTIAL.
 FT CHAIN 27 504 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 113 286 CATALYTIC.
 FT DOMAIN 287 378 PDZ 1.
 FT DOMAIN 401 491 PDZ 2.
 FT ACT_SITE 140 140 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 170 170 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 244 244 CHARGE RELAY SYSTEM (POTENTIAL).
 FT CONFLICT 14 15
 FT CONFLICT 39 147
 FT PSFANVAVDAYSPAVSVVAVQARERVDSDSNFTDFGGRG
 EDIPEDHPLRFPEFAFENDRADRDNRDRGRGRLP
 RAQSGFFINEDGYLTNNHVSIDGSA -> AVSMSTP
 RRRSSPACHVNASATMKATSPISAAAGRTKRTKRCG
 VSSANSLRKVTVPFVIGATIAVRVAKVSVRGRAPASSP
 KTVTSSEPTTSSPTART (IN REF. 1).
 KSADVDLVKVINNAKDKGRSALPQIEAQSGRFVALPTQG
 -> NROTFGR (IN REF. 1).
 FT CONFLICT 464 504
 FT SEQUENCE 504 AA; 5303 MW; D7E82B9981EA23C CRC64;
 Query Match 31.7%; Score 747; DB 1; Length 504;
 Best Local Similarity 36.6%; Pred. No. 5; Le-40;
 Matches 175; Conservative 87; Mismatches 162; Indels 54; Gaps 11;
 10 LAAALCALLA---GCKKAGSFFGADKKEASFVERIETHTKDGSVSMLLPDAQVQSEGP 66
 7 VAGLAVALTLITGIPAEVQASFAEAVRQAAV-----PSFANVADVSP 50
 67 AVVNIOAAPARTQNSGNAETD----SDPLADSPFEFFKLVNMPERPOEAD-- 119
 51 AVVSVAVQARERVDSDSNFTDFGGRGFEDLPEDHPLRFRFRFAR---ENDRADRW 106
 120 -----DGLN---FGSGFIISKNGYILTNTHVAVGMSIKVLLNDKREYAKLIGSD 168
 107 RDRGRGGRGRLPRAQSGSFFITEDGYLTNNHVSIDGSAFTVIMNDGTELDKLVOKD 166
 169 VQSDVALLKIDAEELPVYKIGNPKNLKPEBWAIAIGAPGFDSVYAGVSAKGRSLPN 228
 167 SRDLVLKVDKRRKTTYVSFADEKRVVDWVAVNPFGLGTGTAGTISAGGRDGS 226
 229 ESYTFPIQTDVAINPNSGGPLFNLKQVGINSOIYSRSGFPMGISFATPIDVAMVAE 288
 227 GPYDDVLQYDAANRNKSGSPTFNLSEGVGINTAIRSPSGANGVAFALPASAXKVVVD 286
 289 QLKNTGKVGQGLGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIYLS 348
 287 SLIKDGTVSRGWIQVGIQPTKDIABSLGLSEANGALVVEPQASPEKAGIKNGDVVTA 346
 349 LOGGEIRSSGDLPMVNGAIRPGKEVSLGVWRKEEITIKAKLG---NAAEHTGASSKTD 404
 347 LGEPEPKDPRDLARRVAALRPGSTAEVTLWRSGSEFVNLEIGLTPDAKEPADA---TG 403
 405 EAPYTEQSGSTFESVESAGITLQTHTDSGKGLVVVY---SDAERAGLRHGDIELAV 459
 404 EAGPDGCGAGEALADLGLTIVTSED--GKGVITIASVDPDSADGR-GLKEGETIVEY 458
 RESULT 2
 DEGP BARHE STANDARD; PRT; 503 AA.
 ID _DEGP BARHE
 AC P54925;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable periplasmic serine protease DO-like precursor (EC 3.4.21.-)
 DE (Antigen hcrA).
 CN DEGP OR HTRA.
 OS Bartonella henselae (Rochalimaea henselae).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Bartonellaceae; Bartonella.
 OX NCBI_TaxID=36323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Houston-1;
 RX MEDLINE=94299828; PubMed=8027347;
 RA Anderson B., Sims K., Regnery R., Robinson L., Schmidt M.J.,
 RA Goral S., Hager C., Edwards K.;
 RT "Detection of Rochalimaea henselae DNA in specimens from cat scratch
 disease patients by PCR."
 RL J. Clin. Microbiol. 32:942-948(1994).
 CC - SUBCELLULAR LOCATION: Periplasmic (potential).
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC - SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; L20127; AAA97430.1; -.
 DR MEROPS; S01.273; -.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser. protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 2.
 DR PRINTS; PRO0834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 KW Hydrolyase; Serine protease; Periplasmic; Repeat; Signal.
 FT SIGNAL 1 18 POTENTIAL.
 FT CHAIN 19 503 PROBABLE PERIPLASMIC SERINE PROTEASE DO-
 FT LIKE.
 FT DOMAIN 286 357 PDZ 1.
 FT DOMAIN 419 466 PDZ 2.
 FT ACT_SITE 143 143 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 173 173 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
 FT SEQUENCE 503 AA; 54114 MW; 6CD9F4743282AF9E CRC64;
 Query Match 31.1%; Score 731.5; DB 1; Length 503;
 Best Local Similarity 39.5%; Pred. No. 4; Le-39;
 Matches 183; Conservative 76; Mismatches 173; Indels 31; Gaps 11;
 8 FALAALCALLAGCEKAGSFFGADKKEASFVERIETHTKDGSVSMLLPDAQVQSEGA 67
 15 FSAALFETLFPSSGC---GSSIMTTVAHANSV-----FSSLMOQOGFADIVSQVKA 62
 68 VVNIOAAPARTQNSGNAETDSDPLADSDP----FYEFKLVNMPERPOEAD 120
 63 VVSQVQNSNKKKKKEWFFSDPFTSGFDLPDHPKKEFQDDYNDKESNKL-QSRHL 121
 121 GGLFGSGFIISKNGYILTNTHVAVGMSIKVLLNDKREYAKLIGSDVQSDVALIKIDA 180
 122 RPIAFSGSFFISSDGYITNNHVISDGSYANVLDGHEINAKKLIGTDRTDLAVAKVNE 181
 181 TEELEPVYKIGNPKNLKPEBWAIAIGAPGFDSVYAGVSAKGRSLPNESYTFPIQTDVA 240
 182 KRKFSYVDFGDDSKLRVGDWVAIVAINPFGLTGTGTAGVTSARGRDIQTVYDDFIDAA 241
 241 INPNSGGPLFNLKQVGINSOIYSRSGFPMGISFATPIDVAMVAEGLKNTGKVGQRO 300
 242 VNRNNGSGPTDLNGKVVGVNTAIFSPSGANGVAFALPAATANEVQGLIKGLVQKWA 301
 301 LGVIIQEVSYGLAOSFGLDKASGALIAKILPGSPARAGLQAGDIYLSLDGGEIRSSGL 360
 302 LGVOIQPTKRTISDSIGLKEAKGALITPDKG-PAKAGIKRAGDVIIISNGEKINDYRL 360
 361 FVMVGAITPGKEVSLGVWRKEEITIKAKLGNAEHTGASSXTDEAPYT-EQSGSTFSVE 419


```

Db      361 AKRIANMSPGEIVTVGVKMSGKEENIKVLTDSMPED---ENMKDGSKYSEHNSDETLE 417
Qy      420 SAGITLQTHDSSGKHLVWVRV---SDAERAGLRHDEILAV 459
Db      418 DYGLIVAPSDGLG--LVVTVDPDPSDAADK-GIRGDIYIVV 457

RESULT 3
DEGP BRUAB          STANDARD;          PRT;          513 AA.
ID  DEGP BRUAB
AC  044597;
DT  30-MAY-2000 (Rel. 39, Created)
DT  30-MAY-2000 (Rel. 39, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Probable serine protease do-like precursor (EC 3.4.21.-).
GN  DEGP OR HTRA.
OS  Brucella abortus.
OC  Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
   Brucellaceae; Brucella.
NCBI_TaxID=235;
[1]
SEQUENCE FROM N.A.
RC  STRAIN=2308;
RX  MEDLINE=9516590; PubMed=7861951;
   Tatum F.M., Chevillie N.F., Morfitt D.,
   "Cloning, characterization and construction of htra and htra-like
   mutants of Brucella abortus and their survival in BALB/c mice.";
   Microb. Pathog. 17:23-36 (1994).
CC  -1- SURCELLULAR LOCATION: Periplasmic (Potential).
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC  -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U07352; AAA70164.1; -.
DR  InterPro; IPR001478; PDZ.
DR  InterPro; IPR001940; Protease2C.
DR  InterPro; IPR001254; Ser. protease_Try.
DR  Pfam; PF00089; trypsin; 1.
DR  Pfam; PF00595; PDZ; 2.
DR  PRINTS; PR00834; PROTEASES2C.
DR  SMART; SMO0228; PDZ; 2.
DR  PROSITE; PSS0106; PDZ; 2.
DR  Hydrolase; Serine protease; Repeat; Signal.
FT  SIGNAL          1..25
FT  CHAIN           1..513
FT  DOMAIN          125..299
FT  DOMAIN          300..391
FT  DOMAIN          414..500
FT  ACT_SITE        152..152
FT  ACT_SITE        182..182
FT  ACT_SITE        257..257
SQ  SEQUENCE        513 AA; 53483 MW; 53483 MW; DEICEF1959472806 CRC64;

Query Match          29.4%; Score 692.5; DB 1; Length 513;
Best Local Similarity 37.4%; Pred. No. 14e-36;
Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

Qy      9  ALAALCALLAGCEKAGSFFGADKKEASFVERIEHTKDDGSVMLLP---DFAQLVQSRG 65
Db      14  AAVALLSAL-----AGAFVTVGPICALNEARAR-----AVAVTPPPQAGFADLVKVR 61

Qy      66  PAVNVIAQAPAPRTONGSGN---AETDSPPLADSDPFYFF-----KRLV 107
Db      62  PAVVSVRVKQVQETSNRGPOFGFPGFQDLPDGLKRFDFPFMGEPGDSKSDMRGK 121

```

```

Qy      108  PNMPEIQEERADGGGLNFSSGFIISKNGYIITHTHVAGKSIKYLNDKREYAKLIGS 167
Db      122  ANKRPQCHERP-----VAQSGGFVISEDGYVYVNNHVVSDGAYTVVLDGTGLDGLIGA 177
Qy      168  DVQSDVALAKTADTE-BLPVYKIGNPKNLRKGEVVAIAGAFPGDNSTAGIVSKGSL 226
Db      178  DPTDLVLKINPKRKFVYVAFGDDNKRVRGDDVVAANGVPGAGTVISGIVSRGMDI 237
Qy      227  PNESTYPTFIQTDVAINPNSGGPLFNKGVGVGINSQIYSRSGFMGISPAIPIDAMNV 286
Db      238  GAGPYDDFIOIDAANVKNNGSGPAPFDSLSEGVIGINTAIFSPSGSGVGLAFAPISSTAGV 297
Qy      287  AEDLKNKGKVGQSGQLGYITQEVSYGLAQSPGLDKASGALLAKIIPGSPAERAGIQAQIV 346
Db      298  VDQLIKKGSYERGWIGVQIOPVTYDIAASVGLAEKGAIVASPODDGPAAKAKIVAGDVI 357
Qy      347  LSLDGGEIRSSGGLPVVVGVAITPGEVSLGWR--KGEIITK-AKLGNAAEHTASAKT 403
Db      358  TAVNGETVQDPDRLARKVANIAPGEKALTVWRKNKAEIIVTTAAMPNDKSGSGQSDND 417
Qy      404  DEAPYTEQSGGTFVSAGITLQTHDSSGKHLVWVRV---SDAERAGLRHDEILAV 459
Db      418  NDGGQGE-----TLDSGLTVPSED--GKGVVTVDPDPSDAADR-GIRSGDIYIVV 467

RESULT 4
DEGP CHLMU
ID  DEGP CHLMU          STANDARD;          PRT;          497 AA.
AC  Q9PL57;
DT  16-OCT-2001 (Rel. 40, Created)
DT  16-OCT-2001 (Rel. 40, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  Probable serine protease do-like precursor (EC 3.4.21.-).
GN  DEGP OR HTRA OR TC0210.
OS  Chlamydia muridarum.
OC  Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
NCBI_TaxID=83560;
[1]
SEQUENCE FROM N.A.
RC  STRAIN=Mopn / N49;
RX  MEDLINE=20150255; PubMed=10684935;
   Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
   White O., Hickey E.K., Peterson J., Uterback T., Berry K., Bass S.,
   Ianher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
   Rabin M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
   Eisen J., Fraser C.M.,
   "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia
   pneumoniae AR39.";
   Nucleic Acids Res. 28:1397-1406 (2000).
CC  -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC  -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; AE002288; AAF39082.1; -.
DR  MEROPS; S01.273; -.
DR  TIGR; TC0210; -.
DR  InterPro; IPR001478; PDZ.
DR  InterPro; IPR001940; Protease2C.
DR  InterPro; IPR000126; Ser. protease_V8.
DR  InterPro; IPR001254; Ser. protease_Try.
DR  Pfam; PF00089; trypsin; 1.
DR  Pfam; PF00595; PDZ; 2.
DR  PRINTS; PR00834; PROTEASES2C.
DR  PRINTS; PR00839; V8PROTEASE.
DR  SMART; SMO0228; PDZ; 2.
DR  PROSITE; PSS0106; PDZ; 2.

```

```

KW Hydrolyase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 497
FT DOMAIN 128 289
FT DOMAIN 290 381
FT DOMAIN 394 485
FT ACT SITE 143 143
FT ACT SITE 173 173
FT ACT SITE 247 247
SQ SEQUENCE 497 AA; 53294 MW; 8765F350AC66BBF CRC64;

Query Match
Best Local Similarity 37.4%; Score 683; DB 1; Length 497;
Matches 173; Conservative 77; Mismatches 154; Indels 58; Gaps 11;

QY 28 FGADKKEASFERIEHTKDDGSVM--LLPD---FAQLVSGEPVAVNIOAAPARTQ- 80
DB 22 YSAPKQDSSTGICLAASQSDRELSQEDLKEVSRGSRVAQAATPGVYIENFPKTSQA 81
DB 81 -----NGSGNAETDSDPLADSDPFYEFKRL--VNNPEIPEOEADDDGILNFGSGFIISK 133
DB 82 IASPKNGKGFQENPDPYFND-----EFNNRFGGLPSHREGRRPQORAVR--GTGFIYSE 134
QY 134 NGYILTNTHVAVMGSIKVLNDRREYTKLIGSDVQSDVALLKIDATEELPVVKIGNPK 193
DB 135 DGYYVTNNHHVVEDAGKHVTLHDGQKYAKIIGIDPKTDLAIVIKIQ-KNLPFLTFGNSD 193
QY 194 NLKPGEMVAALGAPFGFNSVTAGIVSAKGRS--LPNSYTPFIOTDVAINNGSGGFLFN 252
DB 194 QLOIGDMSTALGNPFGQATVTVGVISAKGRNQHIVDFEDFIOTDAINNGSGGFLFN 253
QY 253 LKQGVVGINSOIYSRSGFGMGSIFAIPIDVAMNVAEQLKNTGKYORQGLVYIQEVSYGL 312
DB 254 INGQVIGNTVAIVSGSGYIGIGFALPSLMAKRVIDQLISDQVTRGFLGVLQPIDSEL 313
QY 313 AQSFGIDKASGALIAKILPGSPAPRAGLQAGDIVLSLDGGEIRSSGDI PVNWGAITPGKE 372
DB 314 AACYKLEKYVALITVYVKGSPAEKAGLRQEDVIVAVNGKEVESLSALRNALISLMPGTR 373
QY 373 VSLGVARKKEEITIKAKLGNAAEHTGASSTIDEAPYEQ---SGTFVSASAITIQTH 428
DB 374 VVLKVRREGKFI-----ELPVTVTQIPAEADVSAIQKMGVRVQNL 413
QY 429 T-----DSSGKHLVVRVSDAERAGLRHGDDEILAV 459
DB 414 TBEICKKLGASDPTRGIFVYSVEAGSPRAAGVPGQLILAV 455

DUT 5
DEGP CHLTR STANDARD; PRT; 497 AA.
AC P18584; O84830;
DT 01-NOV-1990 (Rel. 16, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-like precursor (EC 3.4.21.-) (59 kDa
DE immunogenic protein) (SK59).
GN DEGP OR HTRA OR CT823.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OC NCBI_Taxid=813;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=Serovar L2;
RX MEDLINE=90337348; PubMed=2379836;
RT "Kahane S., Weinstein Y., Sarov I.;
RT "Cloning, characterization and sequence of a novel 59-kDa protein of
RT Chlamydia trachomatis.";
RL Gene 90:61-67(1990).
RN (2)
RP SEQUENCE FROM N.A.
RC STRAIN=D/UW-3/Cx;
RX MEDLINE=9000809; PubMed=9784136;

```

```

RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
RT Chlamydia trachomatis.";
RT Science 282:754-759(1998).
CC -1 SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1 SIMILARITY: CONTAINS 2 PDZ/DRH DOMAINS.
CC -1 CAUTION: REF.1 SEQUENCE WAS INCORRECT: DUE TO SEQUENCING ERRORS
CC THE AUTHORS TRANSLATED THEIR PUTATIVE 59 kDa IMMUNOGENIC PROTEIN
CC ON THE COMPLEMENTARY STRAND TO THAT OF WHAT SEEMS TO BE THE REAL
CC PROTEIN.
CC -----
CC THE SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DB EMBL; AE001355; AAC68420.1; -.
DB EMBL; M31119; AAA23116.1; -.
DB PDB; 1ZDPA; P18584; -.
DB InterPro; IPR001478; PDZ.
DB InterPro; IPR001940; Protease2C.
DB InterPro; IPR001254; Ser_protease_V8.
DB InterPro; IPR002125; Ser_protease_Try.
DB Pfam; PF00089; trypsin; 1.
DB Pfam; PF00595; PDZ; 2.
DB PRINTS; PR00834; PROTEASE2C.
DB PRINTS; PR00839; V8PROTEASE.
DB SMART; SM00228; PDZ; 2.
DB PROSITE; PS0106; PDZ; 2.
KM Hydrolyase; Serine protease; Repeat; Signal; Antigen;
KM Complete proteome.
FT SIGNAL 1 16
FT CHAIN 17 497
FT DOMAIN 128 289
FT DOMAIN 290 381
FT DOMAIN 394 485
FT ACT SITE 143 143
FT ACT SITE 173 173
FT ACT SITE 247 247
SQ SEQUENCE 497 AA; 53244 MW; 86A5B31B864A38BA CRC64;

Query Match
Best Local Similarity 37.0%; Score 683; DB 1; Length 497;
Matches 171; Conservative 75; Mismatches 158; Indels 58; Gaps 11;

QY 28 FGADKKEASFERIEHTKDDGSVM--LLPD---FAQLVSGEPVAVNIOAAPARTQ- 80
DB 22 YSAPKQDSKADICLAIVSSGDQEVSOEDLKEVSRGSRVAQAATPGVYIENFPKTSQA 81
QY 81 -----NGSGNAETDSDPLADSDPFYEFKRL--VNNPEIPEOEADDDGILNFGSGFIISK 133
DB 82 IASPKNGKGFQENPDPYFND-----EFNNRFGGLPSHRE--QORPQORAVNGTGFIVE 134
QY 134 NGYILTNTHVAVMGSIKVLNDRREYTKLIGSDVQSDVALLKIDATEELPVVKIGNPK 193
DB 135 DGYYVTNNHHVVEDAGKHVTLHDGQKYAKIIGIDPKTDLAIVIKIQ-EKLPLTFGNSD 193
QY 194 NLKPGEMVAALGAPFGFNSVTAGIVSAKGRS--LPNSYTPFIOTDVAINNGSGGFLFN 252
DB 194 QLOIGDMVAIAGNPPGQATVTVGVISAKGRNQHIVDFEDFIOTDAINPNNGSGGFLFN 253
QY 253 LKQGVVGINSOIYSRSGFGMGSIFAIPIDVAMNVAEQLKNTGKYORQGLVYIQEVSYGL 312
DB 254 INGQVIGNTVAIVSGSGYIGIGFALPSLMAKRVIDQLISDQVTRGFLGVLQPIDSEL 313
QY 313 AQSFGIDKASGALIAKILPGSPAPRAGLQAGDIVLSLDGGEIRSSGDI PVNWGAITPGKE 372
DB 314 ATCYKLEKYVALITVYVKGSPAEKAGLRQEDVIVAVNGKEVESLSALRNALISLMPGTR 373

```

QY 373 VSLGWRKKEEITITKAKLGNAAEHTGASSKTDEAPYTEQ-----SGTFVSASGITLQTH 428
 DB 374 VLKLVIREG-----KTIETPTVTVQIPREDGVSALQKMGVRQNI 413
 QY 429 T-----DSSGKLVYVVRVSDAERAGLRHGDILLAV 459
 DB 414 TPEICKKLGLADTRIGILVVAVEAGSPAAAGVAPQQLILAV 455

RESULT 6
 DEGP_CHLPN STANDARD; PRT; 488 AA.
 AC Q926T0; Q9JQD7; Q9K1W4;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DB Probable serine protease do-like precursor (EC 3.4.21.-).
 OS DEGP OR HTRA OR CPN0979 OR CP0877.
 GN Chlamydia pneumoniae (Chlamydia pneumoniae).
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 NCBI_TaxID=83558;
 [1]
 [2]
 [3]
 [4]
 [5]
 [6]
 [7]
 [8]
 [9]
 [10]
 [11]
 [12]
 [13]
 [14]
 [15]
 [16]
 [17]
 [18]
 [19]
 [20]
 [21]
 [22]
 [23]
 [24]
 [25]
 [26]
 [27]
 [28]
 [29]
 [30]
 [31]
 [32]
 [33]
 [34]
 [35]
 [36]
 [37]
 [38]
 [39]
 [40]
 [41]
 [42]
 [43]
 [44]
 [45]
 [46]
 [47]
 [48]
 [49]
 [50]
 [51]
 [52]
 [53]
 [54]
 [55]
 [56]
 [57]
 [58]
 [59]
 [60]
 [61]
 [62]
 [63]
 [64]
 [65]
 [66]
 [67]
 [68]
 [69]
 [70]
 [71]
 [72]
 [73]
 [74]
 [75]
 [76]
 [77]
 [78]
 [79]
 [80]
 [81]
 [82]
 [83]
 [84]
 [85]
 [86]
 [87]
 [88]
 [89]
 [90]
 [91]
 [92]
 [93]
 [94]
 [95]
 [96]
 [97]
 [98]
 [99]
 [100]

DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 KW Hydrolyase; Serine protease; Repeat; Signal; Complete proteome.
 FT SIGNAL 1 20
 FT CHAIN 21 488
 FT DOMAIN 119 280
 FT DOMAIN 281 372
 FT DOMAIN 388 476
 FT ACT_SITE 134 134
 FT ACT_SITE 164 164
 FT ACT_SITE 238 238
 SQ SEQUENCE 488 AA; 52311 MW; 0EB7E0F88BF106F49 CRC64;

Query Match 28.4%; Score 668.5; DB 1; Length 488;
 Best Local Similarity 36.1%; Pred. No. 4.3e-35;
 Matches 169; Conservative 80; Mismatches 170; Indels 49; Gaps 12;

QY 10 LMAALCAGLAGECEKAGSEFFGADKKEASFVERIEHTKDDGSVSMLLPFAQLVQSGPAAV 69
 DB 10 LAVLVGSSLLALPLSGQAVG--KKESS---RVSELQVVLKKEISGFSKATKATPAV 63
 QY 70 NIQAAPAR--TONGSGNAETSDPLADSPFY-EFFKRL--VPMPEIPQ-EEADGGL 123
 DB 64 YIESFPKSAVTHPSPGRRGYPENPF---DYFNDEFFNRPGLPSQREKPKSKAVR--- 117
 QY 124 NFGSGFIISKNGYLTNTHVVAAGMSIKVLNDRREYAKLIGSDVQSVALLKIDATEE 163
 DB 118 --GTGFLVSPGXYVTNNHVEDTGKIHVLDGQKPAIVYGLDPKIDLAVIKI-KSON 174
 QY 184 LPVVKINPKNLKGEWVAAGAPFGFEDNSYTAGIVSAKGRS-LPNESTPTPIQTDVAIN 242
 DB 175 LPVLSFNSDHLKAKGDMAIALGNFGLQATVTVGVISAKNRQHLIADPEDTQIDALIN 234
 QY 243 PGNNGGFLPNKGVGVVINSQIYRSRSGFMGISPAIPIDVAMNVAOLKNTKVRQGLG 302
 DB 235 PGNNGGFLPNIDGQVIGVNTAIVSGSGGYIGFAPISPMNRITDQIRDOVTRGFLG 294
 QY 303 VTIQVSVGLAQSFGDLKASGALAKLIPGSPABRAGIQAQIVYSLGGEIRSSGDLV 362
 DB 295 VTIQPIAEIACKLEKVTALVTVVKGSPADAGAKQEBVITAYVKGEDVLSMPFN 354
 QY 363 MVGAITPEKEVSLGVWRKGEIITIKAKLGNAAEHTGASSKTDEAPYTEQSGTFVSASG 422
 DB 355 AVSLMNPDTRIVLKVAEBGKIELPVYVSAQPKEDGMS-----ALQRVG 398

QY 423 ITLQTHDSSGKHL-----VVVRVSDAERAGLRHGDILLAV 459
 DB 399 IRVQNLTPETAKKLGIAPETKGIILISVEPSVAASSGIAPQQLILAV 446

RESULT 7
 HTOA_HAEIN STANDARD; PRT; 466 AA.
 AC P45129;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DB Probable periplasmic serine protease do/hda-like precursor
 DE (EC 3.4.21.-).
 GN H11259.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_TaxID=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Rd / KW20 / ATCC 51907;
 RX MEDLINE=55350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 Kelleys A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
 McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utechtback T.R., Hanna M.C., Nguyen D.T., Saudak D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm J.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.,
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.",
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: Periplasmic (Potential).
CC -1- SIMILARITY: SEEMS TO BE A INTERMEDIATE FORMS BETWEEN E.COLI HTRA
CC (PROTEASE DO) AND HHOA.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32805; AAC22906.1; -.
CC MEROPS: S01.274; -.
DR TIGR: H11259; -.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001940; Protease2C.
DR InterPro: IPR001254; Ser_protease_Try.
DR Pfam: PF00089; trypsin; 1.
DR Pfam: PF00595; PDZ; 2.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 2.
DR K01 Hydrolyase; Serine protease; Periplasmic; Repeat; Signal;
DR KW Complete proteome.
FT SIGNAL 1 29 POTENTIAL.
FT CHAIN 30 466 PROBABLE PERIPLASMIC SERINE PROTEASE
FT DO/HOA-LIKE.
FT DOMAIN 270 361 PDZ 1.
FT ACT_SITE 367 458 PDZ 2.
FT ACT_SITE 120 120 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 150 150 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 226 226 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 466 AA; 49434 MW; ED050A00047B5851 CRC64;
Query Match 27.5%; Score 648; DB 1; Length 466;
at Local Similarity 37.6%; Pred. No. 8e-34;
cche 166; Conservative 76; Mismatches 163; Indels 36; Gaps 13;
QY 29 GADKEASFEVERIEHTKDGVSMLLPDFAQLVQSEGPVAVNIQAAPARTONGSNAET 88
DB 14 GLSVLSTSVAVH-----AQATLPSFVSEQNSLAPMLEKQ--PAVVTLSVEGKAKV 63
QY 89 DS-DPLADSDP--FYEFF-KRLVPMPEIQEADDDGLNFGSGFTI-SKNGYILNTHV 143
DB 64 DSRSPFLDIDPEEFKFFPGDFPAEOPGREGSKRNFGTL--GSGVIINASKGYVLTNNHV 121
QY 144 VAGMGSIKVLNDKREYTKAKLIGSDVQSDVALKIDATEELPVYKIGNPKMLKGEVAA 203
DB 122 IGGADKITVQLQDGRFFKAKLVGKDESDIALVQLEKPSNLTEIKFASDRLRGDTTVA 181
QY 204 IGAPFGFDSVTAGIVSANKRSLPNES--YTPFIQTDVAINPNSGGPLFNLKQVGVIN 261
DB 182 IGNPFLGQTVTSIGVIALGRSTGSDGTYENIQTDAAVNRGNSGALVNLINGELGIN 241
QY 262 SGIYRSRGSGFMGISAIPIDVANNVABQLKNTGKVGQGLGVIIQEVSYGLAOSFGLDKA 321
DB 242 TAILSPSGGNAGIAFAIPSNQASNLVQQLIEFGQVRRGLIGIKGSELNADLAKAFNVAQ 301
QY 322 SGALAKLIPGAPBAEAGLOAGDIVLSLOGGEIRSSGDLPVMGATIRGEVSLGVWRKG 381
DB 302 QCAFVSEVLPKSAAEKAGLADGITAMNNOKISSFAIRAKIATTGAGKEISLTYLRDG 361

QY 382 EETIKAKLGNAAHHTGASSTIDEAPYTEQSGTFESVSGAGITLQTHDSSGKHLVVVR 441
DB 362 KSHDVKMKL-QADSSQSLSKT-ELPALD-----GATLKDYDAKVGKGEITKX 408
QY 442 ---SDAERAGLRGDELAV 459
DB 409 QPNSLAQR-GLKSGDITLIGI 428
RESULT 8
DEGP_RICCN
ID DEGP_RICCN STANDARD; PRT; 508 AA.
AC Q92JA1;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Probable serine protease do-1like precursor (EC 3.4.21.-).
GN DEGP OR HTRA OR RC0166.
OS Rickettsia conorii.
OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
OC Rickettsiaceae; Rickettsiae; Rickettsia.
OX NCBI_TaxID=781;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Malish 7;
RX MEDLINE=21442074; PubMed=11557893;
RA Ogata H., Audic S., Renesto-Audiffren P., Fournier P.-E., Barbe V.,
RA Samson D., Roux V., Cossart P., Weissenbach J., Claverie J.-M.,
RA Raoult D.,
RT "Mechanisms of evolution in Rickettsia conorii and R. prowazekii";
RL Science 293:2093-2098(2001).
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AE008583; AAL02704.1; ALT_INIT.
DR InterPro: IPR001478; PDZ.
DR InterPro: IPR001254; Ser_protease2C.
DR Pfam: PF00089; trypsin; 1.
DR PRINTS: PR00834; PROTEASES2C.
DR SMART: SM00228; PDZ; 2.
DR PROSITE: PS50106; PDZ; 1.
DR K01 Hydrolyase; Serine protease; Repeat; Signal; Complete proteome.
FT SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 508 PROBABLE SERINE PROTEASE DO-LIKE.
FT DOMAIN 119 284 CATALYTIC.
FT DOMAIN 286 377 PDZ 1.
FT DOMAIN 413 497 PDZ 2.
FT ACT_SITE 134 134 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 164 164 CHARGE RELAY SYSTEM (POTENTIAL).
FT ACT_SITE 242 242 CHARGE RELAY SYSTEM (POTENTIAL).
SQ SEQUENCE 508 AA; 55599 MW; D2F53A690ECDAD7 CRC64;
Query Match 27.1%; Score 638; DB 1; Length 508;
Best Local Similarity 36.8%; Pred. No. 3.8e-33;
Matches 153; Conservative 85; Mismatches 152; Indels 26; Gaps 11;
QY 57 PAQLVQSEGPVAVNIQAAPARTONGSNAETSDPLADSPFYEFFKRLVNMPEIQE 116
DB 56 FADIVEPLIPAVNNISTIEYNSK--SENAE--KDPL--QEKVNDPLEKL--NIP-LNLE 106
QY 117 EADD--GGLNAGSGFIISKNGYILNTHVAVAGMSIKVLNDKREYTKAKLIGSDVQSDVA 174
DB 107 EVDQTPKSPVLSGSGFIIEPNGLIVTNVYIANVDKINIKLADNTELSAKLIGNDITKTDLA 166

Qy	175	LLKIDATEELPVVVKIGDKPNLKPGEWVAIGAPFG- FDNSYTAGIVSAKGRSLPNES---	230
Db	167	LLKIDSEELPVEVEGDSNDARVGMVTAIGNPFGNLGTVISGILSSKGRDIDIDTNI	226
Qy	231	YTPFIQTDVAINPFGNSGGLPFLNKKQVWVINGISQIYSRSRSGFEMGIFSAFPIPIVAMNVAEQL	290
Db	227	VDNFIQTDAAINNNGSGGMFNLDOKVIGVNMVAISPLGTNIGIOFALPSMTAKPIIRRL	286
Qy	291	KNTGVKVGQGLCVITIQEVSVYGLAOSFGIDPKASGALIAKILPGSPAPERAGLOAGDIVSLD	350
Db	287	KDDGVKVSGRGLGVITIQDLETDISEGLKMTGCVLVAKVQEDGFPDCKAGIKGTGDIIFBA	346
Qy	351	GGEIRSSGDLPLMVAVAITPKEVSLGWVRKSGEITIKAKLGNAAHT--GASKTDEAPY	408
Db	347	DIPVNTKKLRYTIADAPIDQEVKVKILDKKLELPILKITSNDNEVTKDSSTETNKKEI	406
Qy	409	TEOOSGTFVSASAGITL-----QTHYDSSGKHLVVVRVSDAERAGLRHDEI	456
Db	407	TNKEBNNSITKNNITTFGNLTBELRKQKTTIPDCKRGIVITNID-EESSFKIGDLI	461

BLAST 9

ID	DEGP	SALTY	STANDARD;	PRT;	475 AA.
DT	01-AUG-1992	(Rel. 23, Created)			
DT	01-AUG-1992	(Rel. 23, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DN	Protease	de precursor (EC 3.4.21.-)			
DE	DEGP	OR HTRA OR PTD OR STM0209.			
OS	Salmonella typhimurium.				
OC	Bacteria; Proteobacteria; gamma sub	division; Enterobacteriaceae;			
OC	Salmonella.				
OX	NCBI_TaxID=602;				
LN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=CS;				
RX	MEDLINE=91251770; PubMed=1645840;				
RA	Johnson K., Charles I., Dougan G., Pickard D., O'Gaora P., Costa G.,				
RA	Ali T., Miller I., Hormaeche C.;				
RT	"The role of a stress-response protein in Salmonella typhimurium				
RT	virulence.";				
RL	Mol. Microbiol. 5:401-407(1991).				
LN	[2]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=LT2 / SGSC1412 / ATCC 700720;				
RX	MEDLINE=21534948; PubMed=11677609;				
RA	McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,				
RA	Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,				
RA	Leonard S., Nguyen C., Scott K., Holmes A., Grewal N.S., Mulvaney E.,				
RA	Ryan R., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,				
RA	Waterston R., Wilson R.K.;				
RT	"Complete genome sequence of Salmonella enterica serovar Typhimurium				
RT	LT2.";				
RL	Nature 413:852-856(2001).				
CC	-1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.				
CC	INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. SHARED				
CC	SPECIFICITY WITH HHOA/DECO				
CC	-1- SUBCELLULAR LOCATION: Periplasmic.				
CC	-1- INDUCTION: BY HEAT SHOCK.				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.				
CC	-1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
OR	EMBL; X54548; CAA38420.1; -.				

Query Match	Similarity	Score	DB	Length
Best Local Similarity	34.18%	Pred. No. 6.7e-33		
Matches 151;	Conservative 83;	Mismatches 156;	Indels 53;	Gaps 9;
DR EMU1; AE008704; AAL19173.1; -				
DR PIR; S15337; S15337				
DR PIR; S21327; S21327				
DR MEROPS; S01.273; -				
DR StyGene; SG10173; degP				
DR InterPro; IPR001478; PDZ				
DR InterPro; IPR001940; Protease2C				
DR InterPro; IPR001254; Ser_protease_TY				
DR Pfam; PF00089; trypsin; 1				
DR Pfam; PF00595; PDZ; 2				
DR PRINTS; PR00834; PROTEASES2C				
DR SMART; SM00228; PDZ; 2				
DR PROSITE; PS50106; PDZ; 2				
KM Hydrolyase; Serine protease; Heat shock; Periplasmic; Repeat; Signal; Complete proteome				
KM Complete proteome				
FT SIGNAL	1	26	BY SIMILARITY	
FT CHAIN	27	475	PROTEASE DO	
FT DOMAIN	281	372	PDZ 1	
FT DOMAIN	378	467	PDZ 2	
FT ACT_SITE	132	132	CHARGE RELAY SYSTEM (POTENTIAL)	
FT ACT_SITE	162	162	CHARGE RELAY SYSTEM (POTENTIAL)	
FT ACT_SITE	237	237	CHARGE RELAY SYSTEM (POTENTIAL)	
SQ SEQUENCE	475 AA;	49315 MW;	86E65BF3C1A289F CRC64;	
Query Match	26.9%	Score 633.5;	DB 1;	Length 475;
Best Local Similarity	34.18%	Pred. No. 6.7e-33		
Matches 151;	Conservative 83;	Mismatches 156;	Indels 53;	Gaps 9;
DR EMU1; AE008704; AAL19173.1; -				
DR PIR; S15337; S15337				
DR PIR; S21327; S21327				
DR MEROPS; S01.273; -				
DR StyGene; SG10173; degP				
DR InterPro; IPR001478; PDZ				
DR InterPro; IPR001940; Protease2C				
DR InterPro; IPR001254; Ser_protease_TY				
DR Pfam; PF00089; trypsin; 1				
DR Pfam; PF00595; PDZ; 2				
DR PRINTS; PR00834; PROTEASES2C				
DR SMART; SM00228; PDZ; 2				
DR PROSITE; PS50106; PDZ; 2				
KM Hydrolyase; Serine protease; Heat shock; Periplasmic; Repeat; Signal; Complete proteome				
KM Complete proteome				
FT SIGNAL	1	26	BY SIMILARITY	
FT CHAIN	27	475	PROTEASE DO	
FT DOMAIN	281	372	PDZ 1	
FT DOMAIN	378	467	PDZ 2	
FT ACT_SITE	132	132	CHARGE RELAY SYSTEM (POTENTIAL)	
FT ACT_SITE	162	162	CHARGE RELAY SYSTEM (POTENTIAL)	
FT ACT_SITE	237	237	CHARGE RELAY SYSTEM (POTENTIAL)	
SQ SEQUENCE	475 AA;	49315 MW;	86E65BF3C1A289F CRC64;	
Query Match	26.9%	Score 633.5;	DB 1;	Length 475;
Best Local Similarity	34.18%	Pred. No. 6.7e-33		
Matches 151;	Conservative 83;	Mismatches 156;	Indels 53;	Gaps 9;
DR EMU1; AE008704; AAL19173.1; -				
DR PIR; S15337; S15337				
DR PIR; S21327; S21327				
DR MEROPS; S01.273; -				
DR StyGene; SG10173; degP				
DR InterPro; IPR001478; PDZ				
DR InterPro; IPR001940; Protease2C				
DR InterPro; IPR001254; Ser_protease_TY				
DR Pfam; PF00089; trypsin; 1				
DR Pfam; PF00595; PDZ; 2				
DR PRINTS; PR00834; PROTEASES2C				
DR SMART; SM00228; PDZ; 2				
DR PROSITE; PS50106; PDZ; 2				
KM Hydrolyase; Serine protease; Heat shock; Periplasmic; Repeat; Signal; Complete proteome				
KM Complete proteome				
FT SIGNAL	1	26	BY SIMILARITY	
FT CHAIN	27	475	PROTEASE DO	
FT DOMAIN	281	372	PDZ 1	
FT DOMAIN	378	467	PDZ 2	
FT ACT_SITE	132	132	CHARGE RELAY SYSTEM (POTENTIAL)	
FT ACT_SITE	162	162	CHARGE RELAY SYSTEM (POTENTIAL)	
FT ACT_SITE	237	237	CHARGE RELAY SYSTEM (POTENTIAL)	
SQ SEQUENCE	475 AA;	49315 MW;	86E65BF3C1A289F CRC64;	
Query Match	26.9%	Score 633.5;	DB 1;	Length 475;
Best Local Similarity	34.18%	Pred. No. 6.7e-33		
Matches 151;	Conservative 83;	Mismatches 156;	Indels 53;	Gaps 9;
DR EMU1; AE008704; AAL19173.1; -				
DR PIR; S15337; S15337				
DR PIR; S21327; S21327				
DR MEROPS; S01.273; -				
DR StyGene; SG10173; degP				
DR InterPro; IPR001478; PDZ				
DR InterPro; IPR001940; Protease2C				
DR InterPro; IPR001254; Ser_protease_TY				
DR Pfam; PF00089; trypsin; 1				
DR Pfam; PF00595				

RX NCBI_TaxID=562;
 RP (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110;
 RA Base S., Gu O., Goddard A.;
 RL Submitted (Oct-1994) to the EMBL/GenBank/DBJ databases.
 RP (2)
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 28-32.
 RC STRAIN=K12 / W3110;
 RX MEDLINE=96165272; PubMed=8576051;
 RA Waller P.R., Sauer R.T.;
 RL "Characterization of degQ and degS, Escherichia coli genes encoding homologs of the Degr protease."
 J. Bacteriol. 178:1146-1153(1996).
 RP (3)
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
 RL "The complete genome sequence of Escherichia coli K-12."
 Science 277:1453-1474(1997).
 CC -1- FUNCTION: PROTEASE WITH A SHARED SPECIFICITY WITH DEGP/DEGP.
 CC -1- SUBCELLULAR LOCATION: Periplasmic.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U15661; AAC33992.1; -;
 DR EMBL; U32495; AAC44005.1; -;
 DR EMBL; U18997; AAA58036.1; -;
 DR EMBL; AB000402; AAC76266.1; -;
 DR MEROPS; S01.274; -;
 DR SWISS-2DPAGE; P39099; COLI.
 DR Ecogene; EG12612; degQ.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 2.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS0106; PDZ; 2.
 DR Hydrolase; Serine protease; Periplasmic; Repeat; Signal;
 KM Complete proteome.
 FT SIGNAL 1 27
 FT CHAIN 28 455 PROTEASE DEGP.
 FT DOMAIN 258 349 PDZ 1.
 FT ACT_SITE 355 447 PDZ 2.
 FT ACT_SITE 109 109 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 139 139 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 214 214 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 455 AA; 47205 MW; 6A090F93AC021C83 CRC64;
 QY Query Match 26.7%; Score 627.5; DB 1; Length 455;
 DB Best Local Similarity 37.2%; Pred. No. 1,5e-32;
 DB Matches 155; Conservative 72; Mismatches 141; Indels 49; Gaps 7;
 QY 54 LPDPAQVQSEGPAAVNIQAPAPRTQNGSGNAETDSDPLADSDPFYEFKRLVPMNPEI 113
 DB 39 LPSLAPMLEKVLPAVAVSVRV-----EGTASQGGKRIEFPFKKFGCDLLPQPAQ 86
 QY 114 POEADDDGGLNGSGFTI-SKNGYIITNTHVAVGMSIKVLLDKRKYAKLIGSDVQSD 172
 DB 114 POEADDDGGLNGSGFTI-SKNGYIITNTHVAVGMSIKVLLDKRKYAKLIGSDVQSD 172

DB 87 PFE-----GLSGGVIIINASKGYVLTNNHVINQAKISIQLDNGREFDPAKLIGSDQSD 139
 QY 173 VALKIDATEELPVYKIGNPKNLKGEWAAIAGAFPGFDSNTAGIVAKGRS-LPNESY 231
 DB 140 IALLQIQNPSTLTQIALINDSKLAVGPRVAVAGNFGIGQRTTSIVSALGSSGLNLEL 199
 QY 232 TPTIQTVAINPNPNSGGPFLNLKGQVGVINSQIYRSRSGGFMGISFPAIPDVANVAEQK 291
 DB 200 ENFIQTDAISIRNGNSGALLNLNGELIGINTAILAPGGSGVIGIPSPNMARTLAQOLI 259
 QY 292 NTGKVGQGGQGVIIQEVSYGLAQSFGDKASGALLAKTLPGSPARAGLQANDIYLSLG 351
 DB 260 DFGIKRGLLTGKTEWSADIAKAFNLDVQGFVSEVLPGSGSAKAGVADIIITSLNG 319
 QY 352 GEIRSGDLPVMVGAITPKEVSLGVMKEGEITIKALGNAAEHTGASSTKDEAPYTEQ 411
 DB 320 KPLNSPFLERSRIRNTTEGTGVYKGLRNGRFLREVLTIST----- 362
 QY 412 QSGTFVESAGITLTQHTDSSG-----KHLVYVVSDAERAGLRHGDILLAV 459
 DB 363 -SSSASAEMITPALEGATLSDQKDGKGKIKIDEVKSGPAAQ-AGLQKDVIIIGV 417
 RESULT 11
 DEGP_RICPR
 ID DEGP_RICPR STANDARD; PRT; 513 AA.
 AC 005942;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable serine protease do-1ike precursor (EC 3.4.21.-).
 GN DEGP OR HTRA OR RP124.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 CC Rickettsiaceae; Rickettsiidae; Rickettsia.
 CC NCBI_TaxID=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=99039499; PubMed=9283893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O., Sichenitz-Ponten T., Aismark U.C.M., Podowski R.M., Naeslund A.K., Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RL "The genome sequence of Rickettsia prowazekii and the origin of mitochondria."
 RT Nature 396:133-140(1998).
 RL Nature 396:133-140(1998).
 RN [2]
 RP SEQUENCE OF 161-513 FROM N.A.
 RC STRAIN=Madrid E;
 RX MEDLINE=97419517; PubMed=9274032;
 RA Andersson J.O., Andersson S.G.E.;
 RL "Genomic rearrangements during evolution of the obligate intracellular parasite Rickettsia prowazekii as inferred from an analysis of 52015 bp nucleotide sequence."
 RT Microbiology 143:2783-2795(1997).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AJ235270; CA14593.1; -;
 DR EMBL; Y11782; CA172471.1; -;
 DR MEROPS; S01.273; -;
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00089; trypsin; 1.

DR Pfam: PF00595; PDZ; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS50106; PDZ; 1.
 KM Hydrolyase; Serine protease; Repeat; signal; Complete proteome.
 FT SIGNAL 1 23 POTENTIAL
 FT CHAIN 24 513 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 124 289 CATALYTIC.
 FT DOMAIN 290 381 PDZ 1.
 FT DOMAIN 418 502 PDZ 2.
 FT ACT_SITE 139 139 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 169 169 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 247 247 CHARGE RELAY SYSTEM (POTENTIAL).
 SQ SEQUENCE 513 AA; 56309 MW; 487E9B7AB4079139 CRC64;
 Query Match 26.2%; Score 615.5; DB 1; Length 513;
 Best Local Similarity 35.4%; Pred. No. 1e-31;
 Matches 149; Conservative 84; Mismatches 157; Indels 31; Gaps 10;
 QY 57 FAQLVQSGPAVNIQAAPAPRTQNGSNAETDSDPLADSPFPEPF-----KRLVP-NM 110
 56 FADIVPELIPAVNI--STIEYVNOKSENKSD--LLQENKHLGFMSPVLEKLANIPML 110
 QY 111 PEIQEADDDGLNGSGFTISKNGYIITNTYVAGMSIKYLANDKREYTAKLIGSDVQ 170
 111 BEIKTPK---SIPKSGFIAPNGILVYVHIANVEKINIKLADNTEFLAKLIGSDSK 167
 QY 171 SDVALLKDATEBELVNVKINPKNLKPEGVAAIGAPFG-FONSVAGIVASGRSLPNE 229
 168 TDLALLKIDSEPLPFVERGDSNDKAVGWVTAIGFNPNGLTGTVSGIISGRDIDVD 227
 QY 230 S---YTPFIQTDVAINPGNSGGLFNLKQVGVNGISQISRSRGFWGIFPAIPDVAMV 286
 228 TDNIYDNIQTDAIINNNGSGPMNLQKVGIVMTAIFSPGTNIGIGFAIPSNNAKPI 287
 QY 287 AEQLKNTKVGQGLGVIIQEVSYGLAQSFGIDKASGALIAKILPGSPAPRAGLQAGDIV 346
 288 IERLKQKQKVRGRGLVITQDLTEBISSEVLGFKGTNGVIVSKQENGPGKAGIKKGDII 347
 QY 347 LSLDGEIRSSGDLFVMVGAITPKEVSLGVWRKGE--RITIKAKLGNAAHNGASKTD 404
 348 IKFGRLVAKNTKGLVITADPIINQSVKKTIRDAQELEPIKVTADNEEVINDSTEETN 407
 QY 405 EAPYEQSGTFSVBSAGITLTQHTDSSGKH-----LVVRYSDAERAGLRHGD 455
 408 KAVIINKKENNLSTIKNNITFENLLELRKXYDIPQDKGIYIINDE--EESVFLGLD 465
 QY 456 I 456
 466 I 466
 RESULT 12
 DEGP_ECOLI STANDARD; PRT; 474 AA.
 ID DEGP_ECOLI AC P09376; P15724;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE DEGP OR HTRA OR PTD OR B0161 OR 20173 OR ECS0165.
 OS Protease do precursor (EC 3.4.21.-).
 OS Escherichia coli, and
 OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 NCBI_taxid=562, 83334;
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=89057448; PubMed=3057417;
 RA Lipinska B., Sharma S., Georgopoulos C.;
 RT "Sequence analysis and regulation of the htra gene of Escherichia
 coli: a sigma 32-independent mechanism of heat-inducible

RT transcription.",
 RL Nucleic Acids Res. 16:10053-10067(1988).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / W3110.
 RX MEDLINE=94261430; PubMed=8202364;
 RA Fujita N., Mori H., Yura T., Ishihama A.;
 RT "Systematic sequencing of the Escherichia coli genome: analysis of
 the 2.4-4.1 min (110,917-193,643 bp) region."
 RL Nucleic Acids Res. 22:1637-1639(1994).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. II, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Schramm S., Duncan M., Allen E., Araujo R., Aparicio A., Chung E.,
 RA Davis K., Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
 RA Lashari D., Lew H., Lin D., Namath A., Oefner P., Roberts D.,
 RA Davis R.W.;
 RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G. II, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimantanta E.T., Pocomusts K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7";
 RL Nature 409:529-533(2001).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC STRAIN=O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasekawa C., Ogasawara N., Yasunaga T.,
 RA Kihara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 O157:H7 and genomic comparison with a laboratory strain K-12."
 RL DNA Res. 8:11-22(2001).
 RN [7]
 RP SEQUENCE OF 1-50 FROM N.A.
 RC STRAIN=K12;
 RX MEDLINE=90323597; PubMed=2165018;
 RA Quirk S., Bhattacharya S.K., Bessman M.J.;
 RT "Primary structure of the deoxyguanosine triphosphate
 triphosphohydrolase-encoding gene (dgt) of Escherichia coli.";
 RL Gene 89:13-18(1990).
 RN [8]
 RP SEQUENCE OF 1-16 FROM N.A.
 RX MEDLINE=90207273; PubMed=2157212;
 RA Wurgler S.M., Richardson C.C.;
 RT "Structure and regulation of the gene for dGTP triphosphohydrolase
 from Escherichia coli."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2740-2744(1990).
 RN [9]
 RP CHARACTERIZATION, AND SEQUENCE OF 27-39.
 RX MEDLINE=90202693; PubMed=2180903;
 RA Lipinska B., Zylitz M., Georgopoulos C.;
 RT "The Htra (DegP) protein, essential for Escherichia coli survival at
 high temperatures, is an endopeptidase."
 RL J. Bacteriol. 172:1791-1797(1990).
 RN [10]

```

RE IDENTIFICATION OF HTRA AND PROTEASE DO
RX MEDLINE=91222240; PubMed=2025286;
RA Seol J.H., Woo S.K., Jung E.M., Yoo S.J., Lee C.S., Kim K.J.,
RA Tanaka K., Ichihara A., Ha D.B., Chung C.H.;
RT "Protease Do is essential for survival of Escherichia coli at high
RT temperatures: its identity with the htra gene product.";
RL Biochem. Biophys. Res. Commun. 176:730-736(1991).
CC -1- FUNCTION: SERINE PROTEASE THAT IS REQUIRED AT HIGH TEMPERATURE.
CC INVOLVED IN THE DEGRADATION OF DAMAGED PROTEINS. IT CAN DEGRADE
CC ICIA, ADA, CASEIN AND GLOBIN. SHARED SPECIFICITY WITH DEGO.
CC -1- SUBUNIT: MULTIMERIC.
CC -1- SUBCELLULAR LOCATION: Periplasmic.
CC -1- INDUCTION: BY HEAT SHOCK.
CC -1- MISCELLANEOUS: HTRA IS INDISPENSABLE FOR BACTERIAL SURVIVAL AT
CC TEMPERATURES ABOVE 42 DEGREES CELSIUS.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC The European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, M36536; AAA23994.1; -.
DR EMBL, X12457; CA030997.1; -.
DR EMBL, D26562; BAA05608.1; -.
DR EMBL, AE000125; AAC73272.1; -.
DR EMBL, U70214; AAB08591.1; -.
DR EMBL, AE005192; AAG54465.1; -.
DR EMBL, AF002550; BAB33588.1; -.
DR EMBL, M29955; AAA23717.1; -.
DR EMBL, M31772; AAA23680.1; -.
DR PIR, S01899; S01899.
DR PIR, B35993; B35993.
DR MEROPS, S01.273; -.
DR SWISS-2DPAGE; P09376; COLI.
DR EcoGene; EG10463; dehp.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00089; trypsin; 1.
DR Pfam; PF00595; PDZ; 2.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
DR PROSITE; PS50106; PDZ; 2.
KW Hydrolyase; Serine protease; Heat shock; Periplasmic; Repeat; Signal;
KW Complete proteome.
FT SIGNAL 1 26
FT CHAIN 27 474
FT DOMAIN 280 371
FT DOMAIN 377 466
FT ACT_SITE 131 131
FT ACT_SITE 161 161
FT ACT_SITE 236 236
FT CONFLICT 10 10
FT CONFLICT 46 46
FT CONFLICT 192 192
FT CONFLICT 467 474
FT SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;
FT PROTEASE DO.
FT PDZ 1.
FT PDZ 2.
FT CHARGE RELAY SYSTEM (POTENTIAL).
FT CHARGE RELAY SYSTEM (POTENTIAL).
FT CHARGE RELAY SYSTEM (POTENTIAL).
FT A -> R (IN REF. 1, 7 AND 8).
FT A -> Q (IN REF. 7).
FT A -> G (IN REF. 7).
FT STYIIMQ -> RHLPNAVVISLNPFLKTGRGSPYNL (IN
FT REF. 1).
SQ SEQUENCE 474 AA; 49354 MW; 5482E596F74B6D5F CRC64;
Query Match 25.9%; Score 610.5; DB 1; Length 474;
Best Local Similarity 34.4%; Pred. NO.1,9e-31;
Matches 146; Conservative 81; Mismatches 149; Indels 49; Gaps 8
OY 54 LPDFACVQVSEGPVNVNIQAAPAPRTQNGSGAETDSP-----LADSPFYEFFKRLV 107
DB 38 MPSLAPMLLEKMPSEVSIINV-----EGSTTVTPPRPRPRFQRFQGDSPFCQGSF- 89
OY 108 PNMPEIIOEEADDDG-----LNFSGFTI-SKNGYILTNTHVAGMSIKVLLNDKREY 160

```

[illegible]

RESULT 13	ID	DEGP_BUCAL	STANDARD;	PRT;	478 AA.
AC	P57322;				
DT	16-OCT-2001	(Rel. 40, Created)			
DT	16-OCT-2001	(Rel. 40, Last sequence update)			
DT	15-JUN-2002	(Rel. 41, Last annotation update)			
DE	Probable serine protease do-like precursor (EC 3.4.21.-) .				
GN	DEGP OR BU228.				
OS	Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum				
OS	symbiotic bacterium) .				
OC	Bacteria; Proteobacteria; gamma subdivision; Buchnera.				
OX	NCBI_TaxID=118059;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=Tokyo 1998;				
RX	MEDLINE=20445173; PubMed=10993077;				
RA	Shigenobu S., Matanabe H., Hattori M., Sakaki Y., Ishikawa H.;				
RT	"Genome sequence of the endocellular bacterial symbiont of aphids				
RL	Buchnera sp. APS.";				
CC	Nature 407:81-86 (2000) .				
CC	-1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.				
CC	-1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch .				
CC	-----				
DR	EMBL; AP001118; BAB12943.1; -.				
DR	MEROPS; S01.273; -.				
DR	InterPro; IPR001478; PDZ.				
DR	InterPro; IPR001940; Protease2C.				
DR	InterPro; IPR001254; Ser.protease_Try.				
DR	Pfam; PF00069; trypsin_1.				
DR	Pfam; PF00535; PDZ; 2.				
DR	PRINTS; PR00834; PROTEASES2C.				
DR	SMART; SM00228; PDZ; 2.				
DR	PROSITE; PS0106; PDZ; 1.				
KE	Hydrolase; Serine protease; Repeat; Signal; Complete proteome.				
FT	SIGNAL	1	26	POTENTIAL.	
FT	CHAIN	27	478	PROBABLE SERINE PROTEASE DO-LIKE.	

FT DOMAIN 116 254 CATALYTIC.
 FT DOMAIN 281 372 PDZ 1.
 FT DOMAIN 387 469 PDZ 2.
 FT ACT_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 228 238 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 478 AA; 52230 MW; 868B8732CAC50629 CRC64;

Query Match 25.3%; Score 594.5; DB 1; Length 478;
 Best Local Similarity 34.1%; Pred. No. 1,9e-30;
 Matches 150; Conservative 86; Mismatches 151; Indels 53; Gaps 10;

QY 43 HTKDDGSVSMILPDPAQLVQSEGPVAVNIQAAPAPRTONGSGNAETDSDFL-----A 94
 DB 31 NTKNSIVREISPSLAPMLEKVPVSI-----NIEGSAITRTSRLPHQFPFG 81
 QY 95 DSDPF-----YEFKRLVPMPEIPOEADDDGLNF---GSGPISKN-CYLLNTH 142
 DB 82 DNSPFCQNSPPRHSPPCHINP-----DSDDKKEKRALGSGVITNADKGYAVTNH 133
 143 VVAGWGIKVLINDKREYAKLIGSDVQSDVALLKIDATEELPVVIGNPKLPGEWVA 202
 134 VVENANKIQVQSDGRYARVIGKDSRSDIALIQKANNLSEITLADSDNLRVGDYTV 193
 QY 203 AIGAPFEDNSVTAGIVSAKGRS-LPNESYTPPIQTDVAINPNSGGLPFLNGQVVGIN 261
 DB 194 AIGNPYGIGETVTSGLISALGRSGNIEHYENFIQTDALINRNGSGALVNLKGELIGIN 253
 QY 262 SQIYRSRGFMKISPAIPIDVAMNVAEQKNTGKVRGQLGVIIQVSYGLASFGIDKA 321
 DB 254 TAILAPDGNIGDIGAIPCNMVKNLTAQVQFQVARGELIGMGEINSDLAQIMKINSQ 313
 322 SGALIAKILPSPARAGLOAGDIYVSLDGEIRSSGDLPVWVGATTPKEVSLGWWRKG 381
 DB 314 KGAFSPRLVPSNSAFEGAKIGADIIISLRKPISSSSSLPAEIGSLPVATKMLGVFRG 373
 QY 382 EBITKAKLGNAAHTGASSKTDEAPYTRQSGTFESVESAGITLQTHTDSGKHLVV--V 439
 DB 374 RIKNITVELKHSVKR-NLNSEND-----YIGIEGVLDSDYIFNEQVIVNDV 420
 QY 440 RVSDAERAGLRHGEITLAV 459
 DB 421 KPHTPASKIGFKKDDIILNV 440

RESULT 14
 DEGP_BUCAP STANDARD; PRT; 478 AA.
 ID DEGP_BUCAP 085291;

16-OCT-2001 (Rel. 40, Created)
 16-OCT-2001 (Rel. 40, Last sequence update)
 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like precursor (EC 3.4.21.-).
 DE DEGP OR HTRA.
 OS Buchnera aphidicola (subsp. Schizaphis graminum).
 OS Bacteria; Proteobacteria; gamma subdivision; Buchnera.
 OX NCBI_TaxID=98794;

RP SEQUENCE FROM N.A.
 RX MEDLINE=98353428; PubMed=9688822;
 RA "Thao M.L., Baumann P.;
 RT "Sequence analysis of a DNA fragment from Buchnera aphidicola (Aphid
 RT endosymbiont) containing the genes dapp-htra-1lvr-1lvr-ftsl-ftsl-
 RT mure";

CC CURT. Microbiol. 37:214-216(1998).
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 2 PDZ/DHR DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb.ch).

DR EMBL; AF060492; AAC3231.1; -
 DR MEMOPS; S01.273; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser-Protease_Try.
 DR Pfam; PF00089; Trypsin_1.
 DR Pfam; PF00595; PDZ_2.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ_2.
 DR PROSITE; PS50106; PDZ; 1.
 DR KEGG; K04462; Serine protease; Repeat; Signal.
 FT SIGNAL 1 26
 FT CHAIN 27 478 PROBABLE SERINE PROTEASE DO-LIKE.
 FT DOMAIN 116 254 CATALYTIC.
 FT DOMAIN 281 372 PDZ 1.
 FT ACT_SITE 133 133 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 163 163 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 228 238 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 478 AA; 51303 MW; C044624F7EFAE988 CRC64;

Query Match 24.3%; Score 571.5; DB 1; Length 478;
 Best Local Similarity 32.4%; Pred. No. 5.4e-29;
 Matches 142; Conservative 81; Mismatches 146; Indels 69; Gaps 9;

QY 53 LLPDPAQLVQSEGPV--VNIQAAPAPRTON-----GSGNAETDSDFLAD 95
 DB 41 LAPSIAEMLEKVPVSIINIEGSTVHTSRLPHQFPFGHNSPFCQNSPFRNSPFCR 100
 QY 96 SDP-----FYEFKRLVPMPEIPOEADDDGLNFGSGPILSKN-CYLLNTHVAGWGI 150
 DB 101 SNRNSNMHEKFNAL-----GSGVITNADKGYAVTNHVENANKI 141
 QY 151 KVLINDKREYAKLIGSDVQSDVALLKIDATEELPVVIGNPKLPGEWVAIAGPFG 210
 DB 142 QVQSDGRYARVIGKDSRSDIALIQKANNLSEITLADSDNLRVGDYTVAILGNPYL 201
 QY 211 DNSVTAGIVSAKGRS-LPNESYTPPIQTDVAINPNSGGLPFLNGQVVGINSQIYRS 269
 DB 202 GEFVTSGLISALGRSGNIEHYENFIQTDALINRNGSGALVNLKGELIGINTALAPDG 261
 QY 270 GFMGISAFAIPIDVAMNVAEQKNTGKVRGQLGVIIQVSYGLASFGIDRASGLIAKI 329
 DB 262 GNIGIGFAIPGNMVKNLTEQVVKRGVKGELIGIEMINSDLAHVMKINQKGFVSGV 321
 QY 330 LPGSPARAGLOAGDIYVSLDGEIRSSGDLPVWVGATTPKEVSLGWWRKEE---IT 385
 DB 322 LPNSSAFAHAGIKAGDIIYSLNKKTISSPALRAEVGSLPVATKMLGIFRNGITGNVIVE 381
 QY 386 IKAKLGNAAE---HTGASSKTDEAPYTRQSGTFESVESAGITLQTHTDSGKHLVVVRV 441
 DB 382 LKESLKNKSVSLGDIYTG-----IEGADLSDCSLNGQGVKIKENTKL 422
 QY 442 SDAERAGLRHGEITLAV 459
 DB 423 NTAQSKIGFKKDDIIVEV 440

RESULT 15
 HTRA_BACSU STANDARD; PRT; 449 AA.
 ID HTRA_BACSU 034358;

DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Probable serine protease do-like htra (EC 3.4.21.-).
 GN HTRA.

OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;

[1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RA Devine K.M.;
 RT "Sequence of the Bacillus subtilis genome between xlyA and ykoR."
 RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Azevedo V., Berto M.G., Beesies P., Boto A., Borchert S.,
 RA Bortiss R., Bortiss R., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Bruch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Comerion I.F., Cummings N.J., Daniel R.A.,
 RA Dentz F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entlan K.D., Errington J., Fabre C., Ferrari E., Folger D.,
 RA Fritz C., Fujita M., Fujita Y., Funa S., Gallizi A., Gallier N.,
 RA Ghim S.Y., Glaeser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaute A.,
 RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karmata D., Kasahara Y., Klaener-Blanchard M., Klein C.,
 RA Kobayashi Y., Koester P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapide A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Paro V., Pohl T.M., Portetelle D., Portollik S., Prescott A.M.,
 RA Preecan E., Puig C., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadle Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi Y., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccori E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpestra P., Tognoni A.,
 RA Toesto V., Uchiyama S., Vandenbol M., Vamier F., Vaasotet A.,
 RA Viati A., Wambut R., Wedler E., Weiler H., Weitzenger T.,
 RA Wintere K., Wipat A., Yamamoto H., Yamane K., Yamamoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Zundin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis".
 RL Nature 390:249-256(1997).
 RN [3]
 RP NATURE 390:249-256(1997).
 RC TRANSCRIPTIONAL REGULATION.
 RA STRAIN=168;
 RX PubMed=10692364;
 RA Noone D., Howell A., Devine K.M.;
 RT "Expression of ykda, encoding a Bacillus subtilis homologue of HtrA,
 RT is heat shock inducible and negatively autoregulated.";
 RL J. Bacteriol. 182:1592-1599(2000).
 RN [4]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX PubMed=11133960;
 RA Noone D., Howell A., Collety R., Devine K.M.;
 RT "ykda and ykta, HtrA-like serine proteases in Bacillus subtilis,
 RT engage in negative autoregulation and reciprocal cross-regulation of
 RT ykda and ykta gene expression.";
 RL J. Bacteriol. 183:654-663(2001).
 RN [5]
 RP TRANSCRIPTIONAL REGULATION.
 RC STRAIN=168;
 RX PubMed=1155295;
 RA Hyvyläinen H.-L., Bolhuis A., Darmon E., Muukkonen L., Koski P.,
 RA Vilkainen M., Sarvas M., Pragai Z., Bron S., van Dijk J.M.,
 RA Kontinen V.P.;
 RT "A novel two-component regulatory system in Bacillus subtilis for the
 RT survival of severe secretion stress.";
 RL Mol. Microbiol. 41:1159-1172(2001).
 CC -1- FUNCTION: May be involved in processing, maturation, or secretion
 CC of extracellular enzymes.
 CC -1- SUBCELLULAR LOCATION: Membrane-bound (Potential).
 CC -1- INDUCTION: Transcription is cases dependent. Induced by heat shock
 CC during exponential growth and by heterologous amylases at the

CC transition phase of the growth cycle. Negatively regulates its own
 CC expression during exponential growth and during heat shock.
 CC -1- MISCELLANEOUS: Inactivation results in compensating overexpression
 CC of ykta, especially during stress conditions.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S2C.
 CC -1- SIMILARITY: CONTAINS 1 PDZ/DRH DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AJ002571; CA005570.1; -
 DR EMBL; Z99110; CAB1347.1; -
 DR MEROPS; S01.273; -
 DR Subtilisin; B012608; htrA.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser protease_Try.
 DR Pfam; PF00089; trypsin; 1.
 DR Pfam; PF00595; PDZ; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 1.
 DR PROSITE; PS0106; PDZ; 1.
 KW Hydrolyase; Protease; Serine protease; Heat shock; Transmembrane;
 KW Complete proteome.
 FT DOMAIN 1 44 CYTOPLASMIC (POTENTIAL).
 FT TRANSEM 45 67 POTENTIAL.
 FT DOMAIN 68 449 EXTRACELLULAR (POTENTIAL).
 FT DOMAIN 103 108 POLY-SER.
 FT DOMAIN 146 152 POLY-SER.
 FT DOMAIN 348 437 PDZ.
 FT ACT_SITE 179 179 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 209 209 CHARGE RELAY SYSTEM (POTENTIAL).
 FT ACT_SITE 290 290 CHARGE RELAY SYSTEM (POTENTIAL).
 SO SEQUENCE 449 AA; 47713 MW; E12B07A9018EE414 CRC64;
 Query Match 21.4%; Score 502.5; DB 1; Length 449;
 Best Local Similarity 35.5%; Pred. No. 1.1e-24;
 Matches 138; Conservative 57; Mismatches 135; Indels 59; Gaps 11;
 QY 30 ADKKEASPERIEHTKDDGVSMLLPFAQLVQSEGPVWVNGAAPPARTONGSNAEDP 89
 DB 98 SEKKSSSSSSAFSEDSKISDMVEDLS-----PAIVGTTNLAQGSNSLFGSSSSD 150
 QY 90 SDPLADSDPFYEFKRLVPMNPEIQEERADGGLNFGSGFIISK-NG--YILTNTHVAG 146
 DB 151 S-----SEDTESGS---GSGVIFKENGKXAYIITNNHVVEG 183
 QY 147 MGSIKVLNDRREYTKLIGSDVQSDVALKIDATEELPVYKIGPKXUKPGEWAATGA 206
 DB 184 ASSIKVSLYDTEVYAKLVGSDSLDLAVLQISDHWTKVNAFNGSSDLRTBEVYAIAD 243
 QY 207 PFGFD--NSVTAGIVSAKGRSL-----ENESYTPRIOTDVAINPNSGGPFLNLKGQVWG 259
 DB 244 PLGKDLSTVTOGIVSGVDRVTSMTSAGETSINVIOTDAALNPGSGGPLNTDCKIVG 303
 QY 260 INSQIYRSRGFMGIFPAIPIDVAMNVNVAEQKNTGKVGOLGVIT-----QEVVS 309
 DB 304 INSMKISBD-DVEGIGFPIPNVDVKPIABELLSKQIRPYIVGVMLEQVPONVQEST 362
 QY 310 YGLAOSFGLDRAAGALIAKIIIPGSPAPAGIQAGDIVSLDGGEIRSSGDL-PVNYGAIT 368
 DB 363 LGL---FGSQLNKGVYIREVVASGSPAEKAGKAEDIIIGLKGKEIDTSGELNNIYKDAK 419
 QY 369 PGKEVSLGVWKKRGREITKALGNAAEHT 397
 DB 420 IGTVEVKILNKGEMTKIKLIDQKEEKT 448

Search completed: July 11, 2003, 10:31:39
Job time : 15 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compen Ltd.

OM protein - protein search, using sw model

Run on: July 11, 2003, 10:25:17 ; Search time 83 Seconds
(without alignments)
1154.360 Million cell updates/sec

Title: US-09-388-090-4
Perfect score: 2353
Sequence: 1 VFKEYQVFAALCAALAG.....EPAGLRHGDDEILAVRASPRQ 465

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

1 number of hits satisfying chosen parameters: 671580

Maximum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_21:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteria:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2246	95.5	499	16	Q9JVT1
2	989.5	42.1	474	2	Q9ALU1
3	985.5	41.9	474	16	Q57155
4	900	38.2	490	16	Q8XPT5
5	896.5	38.1	505	16	Q8Y016
6	894.5	38.0	481	2	Q9AOD1
7	884	37.6	511	2	Q8RTX2
8	862.5	36.7	514	2	Q9PBA3
9	746	31.7	513	16	Q98CS8
10	743	31.6	473	2	Q44476
11	735	30.6	523	16	Q8UGO8
12	719.5	30.6	516	16	Q985F9
13	716	30.4	514	16	Q8UDS7
14	693.5	29.5	513	16	Q8Y332
15	683	29.0	524	16	Q8Y132
16	673	28.6	504	16	Q98KJ1

17	667.5	28.4	403	16	Q8XV99	Q8XV99 ralsionia s
18	664.5	28.2	500	2	Q44652	Q44652 bruceella ab
19	660	28.0	453	2	Q9FD11	Q9FD11 aeromonas h
20	653.5	27.8	457	16	Q8ZB58	Q8ZB58 yersinia pe
21	649	27.6	481	16	Q8ZB56	Q8ZB56 yersinia pe
22	639.5	27.2	478	16	P74978	P74978 yersinia en
23	639	27.2	474	16	Q8YHL4	Q8YHL4 bruceella me
24	637	27.1	437	2	Q68197	Q68197 haemophilus
25	637	27.1	459	16	Q9CMS7	Q9CMS7 pasteurella
26	637	27.1	474	2	Q44596	Q44596 bruceella ab
27	636.5	27.1	475	16	Q8Z9B0	Q8Z9B0 salmonella
28	636	27.0	503	16	Q9Z6C8	Q9Z6C8 rhizobium m
29	636	27.0	530	16	Q9A4S2	Q9A4S2 caulobacter
30	635.5	27.0	477	2	Q8RS81	Q8RS81 klebsiella
31	632.5	26.9	456	16	Q9KJF5	Q9KJF5 vibrio chol
32	632	26.9	463	2	Q68198	Q68198 haemophilus
33	631	26.8	429	16	Q8Y7Z0	Q8Y7Z0 anabaena sp
34	623.5	26.8	455	16	Q8X9F1	Q8X9F1 escherichia
35	621.5	26.4	500	2	Q9KJN6	Q9KJN6 myxococcus
36	617.5	26.2	455	16	Q8Z3E6	Q8Z3E6 salmonella
37	614.5	26.1	459	16	Q9WZ41	Q9WZ41 thermotoga
38	613.5	26.1	455	16	Q8ZTQ1	Q8ZTQ1 salmonella
39	606.5	25.8	452	16	P73354	P73354 synechocyst
40	606.5	25.8	481	16	Q9LEK0	Q9LEK0 shigella so
41	604.5	25.7	491	2	Q9PGL3	Q9PGL3 xylella fas
42	603.5	25.6	371	2	Q31388	Q31388 bradyrhizob
43	600.5	25.5	416	16	P73940	P73940 synechocyst
44	600	25.5	460	2	Q06439	Q06439 rhodobacter
45	589	25.0	389	16	Q9HAX1	Q9HAX1 pseudomonas

ALIGNMENTS

RESULT 1

Q9JVT1 PRELIMINARY; PRT; 499 AA.

AC Q9JVT1
DT 01-OCT-2000 (TREMUREL.15, Created)
DT 01-OCT-2000 (TREMUREL.15, Last sequence update)
DT 01-JUN-2002 (TREMUREL.21, Last annotation update)
DE Putative periplasmic serine protease (EC 3.4.21.).
GN NMA07010.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761915;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holtroyd S., Jagers K., Leather S., Moule S., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrall B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491."
RL Nature 404:502-506(2000).
CC -i- SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
DR EMBL: AL162754; CAB83996.1; -.
DR InterPro: IPR001940; Ser_Protease_Try.
DR InterPro: IPR001478; PDZ.
DR Pfam: PF00595; PDZ, 2.
DR PRINTS: PRO0834; PROTEASES2C.
DR SMART: SM00228; PDZ, 2.
DR PROSITE: PSS0106; PDZ, 1.
DR PROSITE: PSS0240; TRYPSIN_DOM, 1.
KW Hydroxylase; Protease; Serine protease; Complete proteome.
SQ SEQUENCE 499 AA; 52563 MW; 73E979F0A7F5B969 CRC64;

Query Match	Similarity	95.5%	Score 2246	DB 16	Length 499
Best Local Similarity	95.9%	Pred. No. 3	3e-129		
Matches 444	Conservative	6	Mismatches 13	Indels 0	Gaps 0

Query	1	VPKRYQYFALAAACALLAGCEKASFFGADCKEASFYERIEHTTDDGVSMLDPFAOL	60
Db	1	MEKKYQYVLAALCAASLAGCDKASFFGADCKEASFYERIEHTTDDGVSMLDPFVGL	60
Qy	61	VQSESPAVVNIQAAPAPRTONGSGAEFTSDPLADSDPFYEFKRLVPMPEIPOBEAD	120
Db	61	VQSEGPVAVNIQAAPAPRTONGSSNAETDSDPLADSDPFYEFKRLVPMPEIPOBEAD	120
Qy	121	GLMNGSGFTISXNGYILLTNTHVAVGMSIKYLLNDKXEYTKLIGSDVQSVALLKIDA	180
Db	121	GLMNGSGFTISXNDYILLTNTHVAVGMSIKYLLNDKXEYTKLIGSDVQSVALLKIDA	180
Qy	181	TEELPVVNIQGNPKMLPGEWVAIGAPEGFDSVTAGIVASAKRSLPNESYTFIOTDVA	240
Db	181	TEELPVVNIQGNPKMLPGEWVAIGAPEGFDSVTAGIVASAKRSLPNESYTFIOTDVA	240
Qy	241	INPGNSGPPFLNLKQGVVINSQIYSRSGGFEMGISFAIPIDVANNVAEQLKNTGKVRQ	300
Db	241	INPGNSGPPFLNLKQGVVINSQIYSRSGGFEMGISFAIPIDVANNVAEQLKNTGKVRQ	300
Qy	301	LGVIIOEVSYGLAOSFGDLKASGALIAITLPSRPAERAGLQGDIVLSIDGEIRSSGDL	360
Db	301	LGVIIOEVSYGLAOSFGDLKAGALIAITLPSRPAERAGLQGDIVLSIDGEIRSSGDL	360
Qy	361	PVMGAIITPGKEVSLGVWRKKEEITIKKLGNAAHHTGASSTIDEAPYTEQSGTFVS	420
Db	361	PVMGAIITPGKEVSLGVWRKKEEITIKKLGNAAHHTGASSTIDEAPYTEQSGTFVS	420
Qy	421	AGITLQTHTDSSGKRLVVVRVSDAERAAGLRHDEILA VRASP	463
Db	421	AGITLQTHTDSSGCHLVVVRVSDAERAGLRHDEILA VGVVP	463

RESULT 2	Q9ALSI	PRELIMINARY	PRT	474 AA
AC	Q9ALSI			
DT	01-JUN-2001	(Tremblrel. 17, Created)		
DT	01-JUN-2001	(Tremblrel. 17, Last sequence update)		
DT	01-MAR-2002	(Tremblrel. 20, Last annotation update)		
DE	MUCD.			
GN	MUCD.			
PS	Pseudomonas aeruginosa.			
TX	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae; Pseudomonas.			
NCBI	TaxID=287;			
NCBI	(1)			
SEQUENCE FROM N.A.				
RP	STRAIN=UCBPP-PA14;			
RA	Yorgey P.S., Rahme L.G., Tan M., Ausubel F.M.;			
RT	"The Roles of MucD and Alginate in the Virulence of Pseudomonas aeruginosa in Plants, Nematodes, and Mice."			
RL	Submitted (JAN-2001) to the EMBL/Genbank/DBJ databases.			
EMBL	AF343973; AAK1276.1; ..			
InterPro	IPR001478; PDZ.			
InterPro	IPR001940; Protease2C.			
InterPro	IPR001254; Ser_protease_Try.			
DR	Pfam; PF00595; PDZ; 2.			
DR	Pfam; PF00089; trypsin; 1.			
DR	PRINTS; PRO0834; PROTEASES2C.			
DR	SMART; SM00228; PDZ; 2.			
DR	PROSITE; PSS0106; PDZ; 2.			
DR	PROSITE; PSS0240; TRYPsin_DOM; 1.			
KM	Hydrolase; Serine protease.			
SEQUENCE	474 AA; 50349 MW; A22FD4338B859D4C CRC64;			

Query Match	Similarity	42.1%	Score 989.5	DB 2	Length 474
Best Local Similarity	52.5% <td>Pred. No. 1</td> <td>5e-52</td> <td></td> <td></td>	Pred. No. 1	5e-52		
Matches 220	Conservative	55	Mismatches 11	Indels 33	Gaps 11

RESULT 3

OY	54	LPDFAQLOSSEGPAYVNI---QAAAPRPNONGSGNABTDSDDPLADSDP-FYE.	
Dd	27	LPDFTPLVEQASPAVVNISTRKLP----DRAMARQLSTPDLEGLPEPMRDPI	
OY	110	MPEIF---QEADDGGLNGSGFFIISKNGYIIITNTHVAVAGMSIKVLLNDKREYL	
Dd	83	VPRNRPGQQREAO-----SLGSGFIIISNDGYIILTNNHVVADADELIVRLSDRSBHKALVG	136
OY	167	SDVOGDVALLKRTDAIBELPVNATIGNPKNLKPGEWVAIAIGPFPEDNSVTYTGIVSAKRSL	222
Dd	139	ADPRSDVAVLKEA-KNPPTLKLGSNKTKLVGEWVAIGAIPFPFDHSVTYGIVSAKRSL	197
OY	227	PNESTPFPIQTIVA-INPGNSGGPLFNLCGVANGINSIQIYSRSGGFMGISFAIPIDVMNV	286
Dd	198	PNESTVPFIQTIDVA-INPNSGGPLNLBEENVGINOIPIRSSGFMLSTAIPIDIVALNV	257
OY	287	AEOLNKTGRVORGOVLVIIIOEVSYGLAOSFGLDKASKALLAKTLPGSPAEBAIGLOAGDIV	344
Dd	258	ADOLKKAGKVSXGWLGTVIQTENKDLAESFGIDKPSGALVAQIVBDPPAKKGILQGVGDI	311
OY	347	LSLDGGEIRSSDDLVMVGCATITPGKEVSLGWRKGEETITKALGNNAERTGASSKIDE-	407
Dd	318	LSLNGQSINESADLPHLVGNMKRGKINDVIRNCGRKSLSMVAGNLPD-----DDEE	370
OY	406	-----APTQOQSGTFVESYAGITPOTH--TTSSGKHLLVVVRYSDA-AERAGLRHGDEI	456
Dd	371	IASKMAPAERSNSNLTGVTDALTAEQRSLDIG-GVIKYEVDGPAAVITGLRPDVI	428

PRELIMINARY; PRT; 474 AA.

ID	O57155	
AC	O57155;	
DT	01-NOV-1996 (TREMBLrel. 01, Created)	
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)	
DE	MUCD (Serine protease MUCD).	
DN	MUCD OR PA0766.	
OC	Pseudomonas aeruginosa.	
OC	Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;	
OC	Pseudomonas.	
OX	NCBI_TaxID=287;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ATCC 15692 / PAOI;	
RX	MEDLINE=93391358; Pubmed=8378309;	
RA	Martin D.W., Schurr M.J., Mudd W.H., Govan J.R., Holloway B.W.,	
RA	Derecic V.;	
RT	"Mechanism of conversion to mucoidy in Pseudomonas aeruginosa	
RT	infecting cystic fibrosis patients.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 90:8377-8381(1993).	
RN	[2]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ATCC 15692 / PAOI;	
RX	MEDLINE=95286510; Pubmed=7768826;	
RA	Yu H., Schurr M.J., Derecic V.;	
RT	"Functional equivalence of Escherichia coli sigma E and Pseudomonas	
RT	aeruginosa AlgU: E. coli rpoB restores mucoidy and reduces sensitivity	
RT	to reactive oxygen intermediates in algU mutants of P. aeruginosa.";	
RL	J. Bacteriol. 177:3259-3268(1995).	
RN	[3]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=ATCC 15692 / PAOI;	
RX	MEDLINE=96134987; Pubmed=8550474;	
RA	Boucher J.C., Martinez-Salazar J., Schurr M.J., Mudd W.H., Yu H.,	
RA	Derecic V.;	
RT	"Two distinct loci affecting conversion to mucoidy in Pseudomonas	
RT	aeruginosa in cystic fibrosis encode homologs of the serine protease	
RT	HtrA.";	
RL	J. Bacteriol. 178:511-523(1996).	
RN	[4]	
RP	SEQUENCE FROM N.A.	

Query Match	41.9%;	Score 985.5;	DB 16;	Length 474;
Best Local Similarity	52.5%;	Pred. No. 2.7e-52;		
Matches 220; Conservative	55;	Mismatches 111;	Indels 33;	Gaps 11;

RESULT 4	
Q8XPT5	
ID Q8XPT5	PRELIMINARY;
	PRT; 490 AA

DT 01-MAR-2002 (TRENHlrel. 20, Created)
DT 01-MAR-2002 (TRENHlrel. 20, Last sequence update)
DT 01-JUN-2002 (TRENHlrel. 21, Last annotation update)
DE Probable procapsae signal peptide protein (EC 3.4.-.-).
GN R80152 OR R802108
OS *Ralstonia solanacearum* (Pseudomonas solanacearum).
OC plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group.
CC Ralstonia.

Query Match	38.2%;	Score 900;	DB 16;	Length 490;
Best Local Similarity	47.6%;	Pred. No. 4.6e-47;		
Matches 195; Conservative	65;	Mismatches 134;	Indels 16;	Gaps 7;

RESULT 5
Q8Y016
ID Q8Y016 PRELIMINARY; PRT; 505 AA

DT 01-MAR-2002 (TReMBLrel. 20, Created)
 DT 01-MAR-2002 (TReMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Probable periplasmic protease signal peptide protein
 DE (EC 3.4.-.-)
 GN RSC1058 OR RS04156.
 OS *Ralstonia solanacearum* (Pseudomonas solanacearum).
 OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group
 CC Ralstonia.
 OX NCBI_TaxID=305;
 NN [1]

RP SEQUENCE FROM N.A.
 RC STRAIN=GM11000;
 RA MEDLINE=21681879; PubMed=11823852;
 RA Salenouba M., Genin S., Artiguenave F., Gouzy T., Mangenot S.,
 RA Arlet M., Billault A., Broctier P., Camus J.C., Catolico L.,
 RA Chandler M., Choiane N., Claudel-Renard C., Chnacc S., Demange N.,
 RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schlex T.,
 RA Sigler P., Thebaud P., Whalen M., Wincker P., Levy M.,
 RA Weisenbach J., Boucher C.A.;
 RA "Genome sequence of the plant pathogen *Ralstonia solanacearum*,"
 RT Nature 415:497-502(2002).
 DR EMBL; AL646062; CADI4760.1; -
 DR InterPro; IPR001314; Chymotrypsin.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00722; CHYMOTRYPSIN.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 1.
 DR Hydrolyase; Complete proteome.
 SQ SEQUENCE 505 AA; 53008 MW; 07E40468D6DF3142 CRC64;
 Query Match 38.1%; Score 896.5; DB 16; Length 505;
 Best Local Similarity 43.4%; Pred. No. 7.9e-47;
 Matches 207; Conservative 68; Mismatches 147; Indels 55; Gaps 11;
 Oy 8 PALLALCALALGCEAGSGFPGADKKEASFEVERIEHTKDDGVS--SMLLPDPALQVQSE 64
 Db 8 VALRLCVAAVPA--GGALMPAYVGNAT-----GSVATGYGLPDAFADLEVK 54
 Oy 65 GPAVNNIQAPAPRTONGSGNAETDSDPLADSPFEFFFKRL--VP----- 108
 Db 55 SPAVNNIRTEKTKRMNNPSD-----DDMAEFFRRFGVPMPTFGGCGGKRR 103
 Oy 109 NMPEIPEQBEADGGLNFGSGFIIISKNGYILTNTHVVAAGMSIKVLLNDKREYAKLIGSD 168
 Db 104 NMPQAEESQSG--VSGFGIMSGDGYVLTNAHVVEGAETIYVTLIDKKEFAKALIGSD 160
 Oy 169 VQSDVALLKIDATEBELPVYKIGNPVKLKEGEWVAAGAPFGFNSVTAGIVSAKRLPN 228
 Db 161 KRTDVALVVEAT--GLPSLKGDSKRVGEWVLAIGSPGLNTVYAGIVSAKGR--T 217
 Oy 229 ESYTFPIQTDVAINPNSGGLPFLNKGQVVGINSQIYSRSGFMGSAFPIPIVAMVAAE 288
 Db 218 GYLLEFIQSDVAVNPGNSGGLNKGELVGINNQLFSGSGTNGISFALPIDEAMRVAAE 277
 Oy 289 QLKNTGKVGQGLGVIIQEVSYGLAQSGFLDKASGALLAKILPGSPAERAGLQAGDIVLS 348
 Db 278 QLKTGQRTVRGRIIVADIVNPQDAESLGLGRARGAIVGVNESGGPADKAGIEAGDIVLK 337
 Oy 349 LDGGEIRSSGDLPMVGAITTPGKEVSLGVWRKGEETITIRAKLGNAAEHRTASSKTDEAPY 408
 Db 338 FNGRDEKAGDLDQROVGESEKPGFRAIVQWRKATRDLTVTVAELQPDTVAAQKGGQGS 397
 Oy 409 TEOGSGTFVESAGITLQTHDSGKGL-----VVVRVSDA--AEPAGLRHGDELLAV 459
 Db 398 DNGQPGAGKQNALGLVVALDSEGQREFTKAGAVEVQVADGPAARAGIRGDIVILRV 454

OC Pseudomonas.
 ON NCBI_TaxID=321;
 RX [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=FF5;
 RA Keith J.M.W., Bender C.L.;
 RT "Genetic divergence in the *algT*-*muc* operon controlling alginate
 RT biosynthesis and response to environmental stress in *Pseudomonas*
 RT *eryingiae*,"
 RL Submitted (FE8-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF190580; AAK01318.1; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; trypsin; 1.
 DR PRINTS; PR00834; PROTEASES2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 1.
 DR PROSITE; PS50240; TRYPSIN DOM; 1.
 DR Hydrolyase; Serine protease.
 SQ SEQUENCE 481 AA; 50601 MW; E2C9C1137B83920 CRC64;
 Query Match 38.0%; Score 894.5; DB 2; Length 481;
 Best Local Similarity 47.6%; Pred. No. 9.7e-47;
 Matches 202; Conservative 55; Mismatches 128; Indels 39; Gaps 8;
 Oy 54 LPDPAQVQSGPAVNNI---QAPAPRTONGSGNAETDSDPLADSDP--FEFFKRLVP 108
 Db 30 LPDPTGLVQASPAVNNISTRKLPDRAVAN-----QMPDGLCPMLREFLEERSNP 82
 Oy 109 NMPEIPEQBEADG-----GLNFGSGFIIISKNGYILTNTHVVAAGMSIKVLLNDKREYAK 163
 Db 83 PGSPRPGSGAGKGRQREAGSLGSGFIIISPDGYVLTNNHVIDGADELIVRLSDREELKAK 142
 Oy 164 LIGSDVQSDVALLKIDATEBELPVYKIGNPVKLKEGEWVAAGAPFGFNSVTAGIVSAK 223
 Db 143 LVGTDPRTDVAVLKIEG--KDLPTAKLGNSNTLKVEWVLAIGSPGPHSTYKGIIVSAK 201
 Oy 224 RSLPNESTPIQTDVAINPNSGGLPFLNKGQVVGINSQIYSRSGFMGSAFPIPIVDA 283
 Db 202 RSLPNDTVPIQTDVAINPNSGGLPFLNKGQVVGINSQIYSRSGFMGSAFPIPIVDA 261
 Oy 284 MNVAEQKNTGKVGQGLGVIIQEVSYGLAQSGFLDKASGALLAKILPGSPAERAGLQAG 343
 Db 262 MDVANQKANGKAVSRGMLGVITQEVNKLDAESFGLDKPAALVAGLDDGPAKAGGVQ 321
 Oy 344 DIVLSLGGEIRSSGDLPMVGAITTPGKEVSLGVWRKGEETITIRAKLGNAAEH----- 396
 Db 322 DIVLSANGQPIVMSADLPHLIGNLKDGSKALEVLRDCKRQKLTVTGALPDGQEMGDV 381
 Oy 397 --TGASSKTDEAPRTBOGSGTFVESAGITLQTHDSGKGLVYVR--VSDAERAGLRH 452
 Db 382 GGTAA-----ERSSNRLGVSVIELTAQKKSLDLKGVVALKEVTTGCPASLIGLQA 431
 Oy 453 GDEI 456
 Db 432 GDVI 435

RESULT 7
 Q8RTK2
 ID Q8RTK2 PRELIMINARY; PRT; 511 AA.
 AC Q8RTK2;
 DT 01-JUN-2002 (TReMBLrel. 21, Created)
 DT 01-JUN-2002 (TReMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TReMBLrel. 21, Last annotation update)
 DE Protease MucD.
 GN MUCD.
 OS Xanthomonas campestris (pv. campestris).
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 OC Xanthomonas.
 OX NCBI_TaxID=340;

RN [1]
 RP SEQUENCE FROM N.A.
 RA Heu C.-C., Shieh S.-Y., Yang M.-T.;
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.
 SQ EMBL: AF46385; AAL74147.2; -- 368PB263FCB0F713 CRC64;
 SE SEQUENCE 511 AA; 52284 MW; 368PB263FCB0F713 CRC64;
 Query Match 37.6%; Score 884; DB 2; Length 511;
 Best Local Similarity 47.4%; Pred. No. 4.6e-46;
 Matches 203; Conservative 59; Mismatches 118; Indels 48; Gaps 12;
 QY 54 LPDPAQLVQSGPAVAVNIQA-----PADRTONGSGNAETSDPLADSPFFYFFRRL-- 106
 DB 38 LPDFTNLVEQVPGPVNIETTTTRKDAMARSORG--GPGRGCGGAMPDDQMPFFFRFRG 96
 QY 107 -----VPMNP-EIPOEADDDGCL--NFGSGFIISKNGYLLTTHVAVAGMSIVLNDKR 158
 DB 97 PDFQMPGGPGGPGGGDDGGIAGKSGSGFIIISADGYVLTNHHVVDGASEVTVKLTDRR 156
 QY 159 EYTAKLISDVQSDVALLKIDATEELPVYKIGNPKULKEGEMVAIAGAFGPFNDSTAGI 218
 DB 157 EFKAKVQSDGDFVALLKIEA-KGLPTVRIGDSNTLKFGQVVAIGSPFGLDHSTAGI 215
 QY 219 VSAGKRSIP--NESYTPPIQTDVAINPNSGAPLFNLKQVVGINSQIYSRSGGFGISF 276
 DB 216 VSATGRSNPVADQRYVPFIQTVDVAINQNSGAPLNTNRGEVVGINSQITSAGYWGISF 275
 QY 277 AIPIDVAVNAEQLKNTGKVGORGQVITIOEVSYGLAQSGFLDKASGALIAKLIPSPAP 336
 DB 276 AIPIDLASAEQIKASGVNSGMLGVAAGVPIDTLKAOGLGPDTRGALVNDIPAGSPAG 335
 QY 337 RAGLOAGDVLISLDGGEIRSGDLPVNVGAIITPGKEVSLGVWRKGEIITIKALGNAEH 396
 DB 336 KAGIEVGVIVNSVNGKEIAVNASDLPEPMITLMPGTGVSANLVRDGRQVTVTLGLEME 395
 QY 397 TGAASSKTDEAPY-----EQSGSTFVSAGITLQHTTDSGKRLVYRVSD--AAERA-- 448
 DB 396 SGSS-----APRTAADSKSPASAVELLGL-----QVADLTAAERSRN 424
 QY 449 GLRRGDEI 456
 DB 435 GLEAGEGV 442
 RESULT 8
 Q9PBA3 PRELIMINARY; PRT; 514 AA.
 AC Q9PBA3;
 P 01-OCT-2000 (TREMBlrel. 15, Created)
 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Periplasmic protease.
 GN Xf2241.
 OS Xylella fastidiosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
 CC Xylella.
 OX NCBI_TaxID=2371;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=9A5C;
 RX MEDLINE=20365717; PubMed=10910347;
 RA Simpson A.J.G., Reinach F.C., Arruda P., Abreu F.A., Acencio M.,
 RA Alvaranga R., Alves L.M.C., Araya J.B., Bala G.S., Baptista C.S.,
 RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove U.M., Britomes M.R.S.,
 RA Bueno M.R.P., Camargo A.A., Camargo L.E.A., Carraro D.M., Carreir H.,
 RA Colauto N.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
 RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorry H.,
 RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Fetto J.A.,
 RA Fraga J.S., Franca S.C., Franco M.C., Frohme M., Furlan L.R.,
 RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
 RA Ho P.L., Hohenisel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
 RA Krieger J.E., Kurame E.B., Laigret F., Lambais M.R., Leite L.C.C.,
 RA Lemos B.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,

RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marino C.L.,
 RA Marques M.V., Martins E.A.L., Martins E.M.F., Matukuma A.Y.,
 RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
 RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
 RA Nhani A. Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
 RA de Oliveira M.C., de Oliveira R.C., Palmieri D., Paris A.,
 RA Paixoto B.R., Pereira G.A.G., Pereira H.A. Jr., Pasquero J.B.,
 RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
 RA de Rosa V.E. Jr., de Sa R.G., Santelli R.V., Sawasaki H.B.,
 RA da Silva A.C.R., de Silva A.M., da Silva F.R., Silva W.A. Jr.,
 RA da Silveira J.F., Silvestri M.L.Z., Siguelta M.J., de Souza A.A.,
 RA de Souza A.P., Stenzen M.F., Truffi D., Tsai S.M., Tsuchioka M.H.,
 RA Vailada H., Van Sluys M.A., Verjovski-Almeida S., Vettore A.L.,
 RA Zago M.A., Zatz M., Meidanis J., Setubal J.C.;
 RT "The genome sequence of the plant pathogen Xylella fastidiosa";
 RL Nature 406:151-159(2000).
 CC - SIMILARITY: TO SERINE PROTEASES, TRYPSIN FAMILY.
 DR EMBL: AE004037; AAF85040.1; --
 DR InterPro: IPR001478; PDZ.
 DR InterPro: IPR001940; Protease2C.
 DR InterPro: IPR001254; Ser. protease Try.
 DR InterPro: IPR00126; Ser. protease V8.
 DR Pfam: PF00595; PDZ; 1.
 DR Pfam: PF00889; trypsin; 1.
 DR PRINTS: PR00834; PROTEASES2C.
 DR PRINTS: PR00839; V8PROTEASE.
 DR SMART: SM00228; PDZ; 2.
 DR PROSITE: PSS0106; PDZ; 2.
 DR PROSITE: PSS0240; TRYPSIN_DOM; 1.
 DR Hydrolase; Serine protease; Complete proteome.
 SQ SEQUENCE 514 AA; 54140 MW; 707C23FD3C82B8AC CRC64;

Query Match 36.7%; Score 862.5; DB 16; Length 514;
 Best Local Similarity 46.2%; Pred. No. 9.6e-45;
 Matches 194; Conservative 64; Mismatches 129; Indels 33; Gaps 10;
 QY 54 LPDPAQLVQSGPAVAVNIQAAPRTONGSGNAETSDPLADSPFFYFFRRL-----V 107
 DB 54 LPDFTQVQVGVGVNIETTVTRKVGKRGIPLDND-----IPFFFRFFGPDQFM 106
 QY 108 PMPETPOEADDDGCL--NFGSGFIISKNGYLLTTHVAVAGMSIKVLLDKREYTKL 164
 DB 107 PNOGRGQD--DEGIGRAGWGSGFIIISKQYLLTNNHIVIGASGVITKLTDRREFKAKI 164
 QY 165 IGSDVQSDVALLKIDATEELPVYKIGNPKULKEGEMVAIAGAFGPFNDSTAGI 224
 DB 165 IGSDQVQVALLKIDA-KNLPVARIIGDSSSLKSGQVVAIGSPFGLDHSTAGI 223
 QY 225 SLPNES-YTPEFIQTDVAINPNSGAPLFNLKQVVGINSQIYSRSGGFGISFAIPIDVA 283
 DB 224 STSDQRYVPFIQTVDVAINQNSGAPLNTNRGEVVGINSQIFSAGGYMGISFAIPIDVA 283
 QY 284 MNVAEQKNTGYVORGQVITIOEVSYGLAQSGFLDKASGALIAKLIPGSPAERAGI 343
 DB 284 INNAEQIKRTGKQVRSMLGVEIGPIDALKAOGLGLPDSRGALVNNIIPHPSPAAGIEVG 343
 QY 344 DIVLSLDGGEIRSGDLPVNVGAIITPGKEVSLGVWRKGE--EITIKALGNAEHTGASS 401
 DB 344 DVIRSVNGKVISSFSDDLPLIGMPPGTATLGIIRGKPREIVTVLSALN-QDRTASBD 402
 QY 402 KTEDEAPYTEQSGSTFVSAGITLQHTTDSGKRLVYRVSDAERAGLRHGDDELAVRA 461
 DB 403 EDTTAPKERTSA--NVELLGLQVENLSAERRL-----ASSQNAKGVRTITAVTA 452
 RESULT 9
 Q98CS8 PRELIMINARY; PRT; 513 AA.
 AC Q98CS8;
 DT 01-OCT-2001 (TREMBlrel. 18, Created)
 DT 01-OCT-2001 (TREMBlrel. 18, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Serine protease, Htra/Deqo/degS family.

GN ML5022
 OS Rhizobium loti (Mesorhizobium loti).
 CC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 CC Phyllobacteriaceae; Mesorhizobium.
 OX NCBI_Taxid:381;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MAFF303099;
 RC MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Maranda A., Ideawa K., Ishikawa A., Kawashima K., Kiyuna T.,
 RA Kishida Y., Kiyokawa C., Kohara S., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpō S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti";
 RL DNA Res. 7:331-338(2000).
 DR EMBL; AP003005; BABS1543.1; -.
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; Trypsin; 1.
 DR PRINTS; PR00834; PROTEASE2C.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS5106; PDZ; 2.
 KW Protease; Complete proteome.
 QO SEQUENCE 513 AA; 52258 MW; 901FAF72004FE53B CRC64;

Query Match	31.7%	Score 746;	DB 16;	Length 533;
Best Local Similarity	39.0%	Pred. No. 1.2e-37;		
Matches 184;	Conservative 81;	Mismatches 177;	Indels 30;	Gaps 11;

QY	10	LAALCAALILACCENKASFPFGADKKKEASVEVEIETHKDDGSVSMLLPD--FQOLVQSECPA	67
	11	: : : : : : : : : : : : : : :	
Db	12	VALLGALLIISPVVSE--AQSANSTSESVKIYATTOCPVPAIGITRAPNSCFAPYAALDKPA	69
QY	68	VNIQAAPAPRTQNGSNAETDSDPLADSDPEFEFFKRLVENM---PEIQEADOGGL	123
Db	70	VVVTYS-----IMKGPASDDGMPPLGNS-PPDQYFRQPFQDGMAPQTPPOOQARA	123
QY	124	NFGSGFIISKNGYLLTNTHTVVAAGMSIVLVNDKREYAKLIGSVQSDVALLKIDATEE	183
	124	: : : : : : : : : : : : : :	
Db	124	LGSGFVITAGCTVATVNTNNHVVDGASSIKVTLDDGTLEPAKLVGRARKNDLVLKIKSKP	182
QY	184	LPVYKIGNPKMLKGEWEVAAIGAPFGFONSYTAGISAKGRSLPNESYTPETQTDVAINP	243
	183	: : : : : : : : : : : : : :	
	183	LPYTKMGSDRLMTMGDQVLAIGNPPGICTYTTAGIVSARGNDLHSGPDDPDIQIDAPINH	242
QY	244	GNSGGPLFNLKGOVYVINSQIYSRSGFMGISFAPIDVAMNVAAEQLKNTKGVQGRQLGV	303
	243	: : : : : : : : : : : : : :	
Db	243	GNSGGPLVDVNGNVVGINTAIYISPNGGSGVGCFAPISPOQAKVVAKLKMKDSDSIQYGYLGV	302
QY	304	IIQEVSVGLAOSFGIDKASGALLIKILPESPAEBRAGIADGDIYVSLDGGELIRSSGDLPLVM	363
	303	: : : : : : : : : : : : : :	
Db	303	EIOEVTTPVVASIIGLDHAGGALVSKVWSDSPAAASHGVAGAVITTFPAQDVADPKDLSRA	362
QY	364	VGAITPCKEVSIGVWRKGEIITIKAKIG-NAAEHTGASSKTDDEAPYTEQOOSGTFVESAG	422
	363	: : : : : : : : : : : : : :	
Db	363	VADVAPGAKETLIDVWRKCKAMQISVEVQNSDDVKTASAGESGAPSAEQS--RAPAI	419
QY	423	ITLTQTHD-----SSGKI-LVVVRS--DAAERAGLRHGDDELAVRAP	463
Db	420	LGIMDITPDIROENMLAGNEHGAVVARNPDKAAAAGIIOGDDIIVANQAP	471

RESULT	10		
Q44476			
ID	Q44476	PRELIMINARY;	PRT; 473 AA.
AC	Q44476.		
DT	01-NOV-1996	(TRENMBLrel. 01,	Created)
DT	01-NOV-1996	(TRENMBLrel. 01,	Last sequence update)
DT	01-JUN-2002	(TRENMBLrel. 21,	Last annotation update)

```

DE      MucD.
OC      Azotobacter vinelandii.
OC      Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC      Azotobacter.
OX      NCBI_TaxID=354;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=UW 136;
RX      MEDLINE=96178940; Pubmed=8606151;
RA      Martinez-Salazar J.M., Moreno S., Najera R., Boucher J.C., Espin G.,
RT      Soberon-Chavez G., Detetic V.;
RT      "Characterization of the genes coding for the putative sigma factor
RT      AbgU and its regulators MucA, MucB, MucC, and MucD in Azotobacter
RT      vinelandii and evaluation of their roles in alginate biosynthesis.";
RL      J. Bacteriol. 178:1800-1808(1996).
DR      EMBL; U30799; AAB01513.1; -.
DR      InterPro; IPR001478; PDZ.
DR      InterPro; IPR001254; Ser_protease_Try.
DR      Pfam; PF00595; PDZ; 2.
DR      Pfam; PF00089; trypsin; 2.
DR      SMART; SM00228; PDZ; 2.
DR      PROSITE; PSS0106; PDZ; 2.
DR      PROSITE; PSS0240; TRYPSIN_DOM; 1.
KW      Hydrolyase; Serine protease.
SQ      SEQUENCE 473 AA; 50282 MW; C97B357D897738AB CRC64;

Query March 31.6%; Score 743; DB 2; Length 473;
Best Local Similarity 42.8%; Pred. No. 1.7e-37;
Matches 182; Conservative 56; Mismatches 159; Indels 26; Gaps 8

```

50 VSMILPPDPAQVQSEGPVMT---QAAPAPRONGSGNAETSDPLADSDPFVEFFKRL 106

```

Qy 107 VPMNPEI:POEADGGINFGSGFI:ISKGYIL:LTNTHV:VAGMS:IKYLLNDKREY:TAKLIG 166
Db 79 I:POL:PRTPDNG:ROREAH:SLGSGFI:ISPDGYL:TNHVA:DADEI:IVR:SDS:SELEAE:LVG 138
Qy 167 SDVQSDVALLKIDATEELPVYK:GNPNLKGEMVAALIGAPGEHDSNV:TAGIVSAKGSL 226
Db 139 ADPLTRCSFVEGGCFES-PSHQTYTPRIEIRGMSGCHRPFRFRSPDCHCHQCHAKP 197
Qy 227 PNESTYPIOTDVAINP:GNSG:PLFNLKGVAVINSOIYRSRGGFMG:ISFAIPIDVAMNV 286
Db 198 AERLVAFFIOTDVAINP:GNSG:PLFDDKGVITGINSQIFTRSGGFMG:ISFAIPLEVAMGV 257
Qy 287 AEOUKONTGKVORQGLGYIIQEVSYGLAQSPGLDKASGALIAKILPGSAPERAGIOAGDIV 346
Db 258 ADQKATGKVARGLGYIIQEVNMDLAE:SGDLDRPAALVAQVLEDDGADKGLQVGVVI 317
Qy 347 LSLDGEIRSSGDL:PVNVGAITPQKEVSLGVWRKGE---ITTKA--KLGNALEHTGAS 400
Db 318 LSLDGHPIVMSADLPHLVGKLPKPAANLELVWRDGRKRNIATITGALPEEGNGVQPSIAG 377
Qy 401 SKIDEAFTYEQSGSTFEVESAAGITLQHTTSSGHALVYVR--VSDAABRAGLHRGDELLA 458
Db 378 -----TEOSSNRLGVTVELTAEOKKSIDLKGVAVIREVLINGPALIGLRPDEVYTH 429
Qy 459 VRASP 463
Db 430 LNNOP 434

```

RESULT 11		
08UG08		
ID 08UG08	PRELIMINARY;	PRT; 523 AA.
AC 08UG08;		
DT 01-JUN-2002	(TREMBlrel. 21, Created)	
DT 01-JUN-2002	(TREMBlrel. 21, Last sequence update)	
DT 01-JUN-2002	(TREMBlrel. 21, Last annotation update)	
DE	Serine protease DO-like protease.	
DD	DOP OR ATU0977 OR AGR_C_1792.	
CS	Agrobacterium tumefaciens (strain C58 / ATCC 33970).	

OC Bacteria, Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 RN NCBI_TaxID=176299;
 RN [1]
 RX SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Secubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F. Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Bisen J.A., Karp P.D., Boyee D. Sr.,
 RA Chapman P., Cleendenig J., Deatherage G., Giliet W., Grant C.,
 RA Kutayin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Semphimachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Rao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Neeter E.W.;
 RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
 CS8.";
 RT Science 294:2317-2323 (2001).
 RL [2]
 RN SEQUENCE FROM N.A.
 RN MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlel K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
 RA Wolan C., Allinger M., Doughy D., Scott C., Lappas C., Markelz B.,
 RA Planagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens CS8.";
 RL Science 294:2323-2328 (2001).
 RN EMBL; AB009062; AAL4191.1; -
 RN EMBL; AB008028; AAK6786.1; -
 RN DR
 RN Protease; Complete proteome.
 RN SEQUENCE 523 AA; 54436 MW; AFA52B8A5E258ED CRC64;
 SQ
 Query Match 31.2%; Score 725; DB 16; Length 523;
 Best Local Similarity 37.0%; Pred. No. 66-37;
 Matches 168; Conservative 82; Mismatches 154; Indels 50; Gaps 9;
 QY 36 SFVERIETKDDGSVSMILPFAQVYOSBGPVAVNIQAAPARTONGSNAETDS----- 90
 DB 44 SFAAPVETAPQ-----VPSPFANVADVSPVAVSV-----RVGSONVPASDSSNFSG 91
 QY 91 -----DPLADSDPFYEFKRLVNMPEIPOEABDDGLN-----FGSGPI 130
 DB 92 NFGGRGLDQLPDHDLKRFKFEFGQ-----NQDSRGPNNHRODGPRLRPAQSGSFF 146
 QY 131 ISKNGYILTNTHVAVAGMSIKVLNDRREYAKLIGSDVQSDVALIKIDATEBELPVVKIG 190
 DB 147 ISEBGVYVTVNNHVVDDGSAVYVNMVDTGLEAKLVGRDPRDLALKLKVNKRFTYVKA 206
 QY 191 NPKULKGEWVAALGAPPGFNSVTAGIVSAKGRSLPNESYTPFIQTDAVNPNGSGPL 250
 DB 207 DDTIRIVDDWVAVAVPGLGGLTVISGIIISARGRDIGSGPYDDYQIIDAANRNSGGBA 266
 QY 251 FNLGGOVVGINSQIISRSRGFWGISFAIPIDVAVNVAEOLKNTKQVQROQLGVIIQSEVY 310
 DB 267 FNLGSEVVGINTALFSPSGGNVGIAPALPSSVAKVADLVQDQKVERKMLGVQIQPVSK 326
 QY 311 GLAOSFGIDKASGALIAKILPGSPAERAGLQAGDIVLSLDSGEIRSSGDLPMVVGAIITPG 370
 DB 327 DIASLGLAEKAGLIVSPQSGSPGDKGIGKGGDIITANQDPVADARLSRIGMAIN 386
 QY 371 KEVSLGVWRKBEETIKAKLGN--AAERTGASKTDEADYEQSGTFVESAGITLQTH 428
 DB 387 SKYVLSIMRGKSGSVTVTLGDLTDEASKATPQSDNDKSGSGSSEKVLSSGLITVSPS 446
 QY 429 TDSGKHLVVVRV---SDAABRAGLRHGDILLAV 459
 DB 447 DDGNG--LAITVDVDDSDPAAR-GLTKGKITSV 477

RESULT 12
 Q985F9
 ID Q985F9 PRELIMINARY; PRT; 516 AA.
 AC Q985F9;
 DT 01-OCT-2001 (TEMBLrel. 18, Created)
 DT 01-OCT-2001 (TEMBLrel. 18, Last sequence update)
 DT 01-MAR-2002 (TEMBLrel. 20, Last annotation update)
 DE Serine protease.
 GN M17692.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria, Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Pnyliobacteriaceae; Mesorhizobium.
 RN NCBI_TaxID=381;
 RN [1]
 RN SEQUENCE FROM N.A.
 RN STRAIN=MAF303099;
 RX MEDLINE=21082930; PubMed=11214968;
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Saeamoto S.,
 RA Matanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 RT Mesorhizobium loti.";
 RL DNA Res. 7:331-338 (2000).
 DR EMBL; AP003012; BAB54103.1; -
 DR MEROPS; S01.273; -
 DR InterPro; IPR001478; PDZ.
 DR InterPro; IPR001940; Protease2C.
 DR InterPro; IPR001254; Ser_protease_Try.
 DR InterPro; IPR000126; Ser_proteas_v8.
 DR Pfam; PF00595; PDZ; 2.
 DR Pfam; PF00089; tryptsin; 1.
 DR PRINTS; PRO0834; PROTEASE2C.
 DR PRINTS; PRO0839; V8PROTEASE.
 DR SMART; SM00228; PDZ; 2.
 DR PROSITE; PS50106; PDZ; 2.
 DR DR
 RN Protease; Complete proteome.
 RN SEQUENCE 516 AA; 53704 MW; FE8DAADC099ABDC6 CRC64;
 SQ
 Query Match 30.6%; Score 719.5; DB 16; Length 516;
 Best Local Similarity 36.2%; Pred. No. 5.2e-36;
 Matches 174; Conservative 80; Mismatches 177; Indels 49; Gaps 9;
 QY 10 LALCALALAGCKKAGSFFADKKEASFVERIETKDDGSVSMILPFAQVYOSBGPVAV 69
 DB 16 LAAAVAVAVAGVIGVAGLISGTSPLADAVRVEAVQVQ-----FADVERVSPAV 67
 QY 70 N--IQAPAPRTONGSNAETSDPLADSDPFYEFKRLVNMPEIPOEABDDGLN--- 124
 DB 68 SVAVKAKIQPTADGSDQ--QDGFNLPNNQLRFFKEF-----RGFDGGQNDG 118
 QY 125 -----FGSGFIISKNGYILTNTHVAVAGMSIKVLNDRREYAKLIG 166
 DB 119 HRRFGRHRSNDQPRPAVQSGSFFISBDGYLVNTHVVEEGTAVTJNDKEIDAKLVG 178
 QY 167 SDVQSDVALIKIDATEBELPVVKIGNPKULRGEWVAALGAPFGDNSTAGIVSAKGS 226
 DB 179 TDRITLAVLAKVGGGFFTYVDPADDSKVGWVAVVGNPFGIGVTVTAVIARGDI 238
 QY 227 PNESTYTPFIQTDAVNPNGSGPLFNLKGOVVGINSQIISRSRGFWGISFAIPIDVAVN 286
 DB 239 GAGPYDDFLQIDASVNRKNSGGFTFNLNGVVGINTALFSPSGSSVGIAPIPASTACV 298
 QY 287 AEQLKNTGKQVQROQLGVIIQSEVYGLAQSFGIDKASGALIAKILPGSPAERAGLQAGDIV 346
 DB 299 VEDLMKSGAVQRGWLGVEIQPTSDIASISGLKSNNGALVSSAODDGPGRKAGITAGDVI 358
 QY 347 LSLDSGEIRSSGDLPMVVGAIITPGKEVSLGWWRKGEIITIKAKLGN--AAERTGASKTDE 405
 DB 359 TVEGKDVAVSPKELARLIGAVYSPKSVDTVWVRGKQITIKVDIGKLPASDKQASNDQ 418
 QY 406 APYTEQSGTFVESAGITLQTHTDSGKHLVVVRV---SDAABRAGLRHGDILLAVRAS 462

Db 419 QPAAPAKPDTL-----ADLGLVTYTKSENGKGLVYTDVPESAAADR-GIQDPDITTAANSN 473

RESULT 13

08UDS7 PRELIMINARY; PRT; 514 AA.

AC 08UDS7; 01-JUN-2002 (TREMBlrel. 21, Created)

DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)

DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Serine protease DO-like

GN HTRA OR ATU2043 OR AGR_C_3700.

OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Rhizobiaceae; Rhizobium.

OK NCBI_TaxID=176299;

RN [1]

SEQUENCE FROM N.A.

MEDLINE=21608550; PubMed=11743193;

Wood D.W., Secubal J.C., Kaul R., Monke D.E., Kitejima J.F., Okura V.K., Zhou Y., Chen L., Wood G.E., Almeida N.F.Jr., Woo L., Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr., Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C., Kutyavin T., Levy R., Li M.-J., McClelland E., Palmeri A., Raymond C., Rouse G., Saenphimmachak C., Wu Z., Romero P., Gordon D., Zhang S., Yoo H., Tao Y., Biddell P., Jung M., Krespan W., Perry M., Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M., Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V., Neeter E.W.;

"The genome of the natural genetic engineer Agrobacterium tumefaciens C58.";

Science 294:2317-2323(2001).

RN [2]

SEQUENCE FROM N.A.

RP MEDLINE=21608551; PubMed=11743194;

Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M., Quicilo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L., Houtell K., Gordon J., Vaudin M., Iarchouk O., Epp A., Liu F., Mollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B., Flanagan C., Crowell C., Gursen J., Lomo C., Sear C., Strub G., Cielo C., Slater S.;

"Genome sequence of the plant pathogen and biotechnology agent Agrobacterium tumefaciens C58.";

Science 294:2323-2328(2001).

RL EMBL; AB009155; AAL43035.1; -

EMBL; AB008120; AAK87796.1; ALT_INIT.

Protease; Complete proteome.

SEQUENCE 514 AA; 5351 MW; 1138B132E8459EA7 CRC64;

Query Match 30.4%; Score 716; DB 16; Length 514;

Best Local Similarity 36.4%; Pred. No. 8.4e-36;

Matches 170; Conservative 75; Mismatches 146; Indels 76; Gaps 10;

Qy 48 GSVMLLPFAQLVQSEGP-----AVNTIQAPAPRTONGSGNAETDSPLAD 95

Db 24 GSLT---AFVAAARQSHGPESSVADLAEPRLDVAVNISTSQNVKTE-GKGPV---PPKLP 76

Qy 96 SPPFEEFFKRLVPMMEIQOEADDDG---LNGSGFIISKNGVYLLNTHTVVAAGMSIKY 152

Db 77 GSPFOEFFK-----DYPSQKPEEGEKKVNSLGSQFVIDPAGVVTNNHVIKADIEV 129

Qy 153 LNDKREYAKLIGSDVQSDVALKIDATEELPVVTKIGNPKNLKPGEWVAAGAPFGFDN 212

Db 130 IFPNQSKLAKATLVGTDTKTDLSTLTKPEKPLKAVFPGDSRMRIKGMVAAGNPFELGG 169

Qy 213 SYTAGIVSAKSGSLPNESTYTPPIQTQDVAINPNSGGPFLNLKQGVVNGINSQIYSRSGFM 272

Db 190 SLTVIVISARGNHINAGPYDNFIQTDAANKNGSGPFLFMKGEVIGINTAIISPSGSI 249

Qy 273 GISFAIPIDVANNVNAEQLKNTGKVGKQGVIIQEVSYGLAOSFGIDKASGALIAKILPG 332

Db 250 GIGFAVPELAAQNIIVQOLLIEFGETRGMIGRVQPVTDVVAASLGMDSAKALISGAVK 309

Qy 333 SPAERAGIQADIVLSLDGGEIRSSGDI.PVVYGAITPCKEVS LGWRKGEITITIAKLG 392

Db 310 GPVENGPQAGDVVLKFKDGKIDINEMRDLIRIVASPVKEVDVVYRRGKEETVKKVGLGQ 369

Qy 393 AAHTTGASSKTDEAPYREQSG-----TFVSASAGITLQTHDSSK 434

Db 370 LQDTTDEKASTD--PGEEDGGSWVAPDDKDGDDAQDQTFPEVEAPQTV-----LGM 422

Qy 435 HLVV-----VRSPDAERAGRHGDEILAV 459

Db 423 NLVVLNSBELREKGIASVEGVVAVSDPGSPALQKKGAKADVIVEV 469

RESULT 14

08YG32 PRELIMINARY; PRT; 513 AA.

AC 08YG32; 01-MAR-2002 (TREMBlrel. 20, Created)

DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)

DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)

DE Protease DO (EC 3.4.21.-).

GN BME11330

OS Brucella melitensis.

OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;

OC Brucellaceae; Brucella.

OK NCBI_TaxID=29459;

RN [1]

SEQUENCE FROM N.A.

RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;

RL MEDLINE=20020109; PubMed=11756688;

RA Delvecchio V.G., Kapatal V., Redkar R.J., Patra G., Muijer C., Los T., Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G., Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Gottsmann E., Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-J., Haeselorn R., Kyrides N., Overbeek R.;

"The genome sequence of the facultative intracellular pathogen Brucella melitensis.";

Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).

DR EMBL; AB009571; AAL52511.1; -

DR InterPro; IPR001478; PDZ.

DR InterPro; IPR001940; Protease2C.

DR InterPro; IPR01254; Ser_protease_Try.

DR Pfam; PF00595; PDZ_2.

DR Pfam; PF00089; trypsin_1.

DR PRINTS; PRO0834; PROTEASES2C.

DR SMART; SM00228; PDZ_2.

DR PROSITE; PSS0106; PDZ_2.

KW Hydroxylase; Complete proteome.

SEQUENCE 513 AA; 5351 MW; C472FEF99DFC6268 CRC64;

Query Match 29.5%; Score 693.5; DB 16; Length 513;

Best Local Similarity 37.4%; Pred. No. 2e-34;

Matches 179; Conservative 82; Mismatches 165; Indels 53; Gaps 13;

Qy 9 ALAALCAALLAGCEKAGSFFGADKKKASFYERIKHTKDDGVSMLLP---FAQLVQSEGS 65

Db 14 AAVVLSAAL-----AGAFVVTGPLGALNEARAE-----AVHTVPPQAGAFADLVEKVR 61

Qy 66 PAVVNTQAAPAPRTONGSGN---AETDSPLADSDPFEFF-----KRLV 107

Db 62 PAVVSVKVDVQVQTSNRKGPQFGPPGPDQLPDGHPLEKRFPRDFGMEPRGDSRSDNRK 121

Qy 108 PNMPEIQOEADDDGLNFGSGFIISKNGVYLLNTHTVVAAGMSIKYLNDRKREYAKLIGS 167

Db 122 ANKRPQHERP---VAQSGSFVISEDGVVTNNHVSDDGAIYVLDGTELDKALIGA 177

Qy 168 DVQSDVALKIDATE-ELPVVTKIGNPKNLKPGEWVAAGAPFGDINSYTAGIVSAKGRSL 226

Db 178 DPTDVLAVLKINAQRKFVYVAFGDDNKNRVGDMVVAAGNPFGLGCTVTSIGVARGRDI 237

Qy 227 PNESTYTPPIQTQDVAINPNSGGPFLNLKQGVVNGINSQIYSRSGFMGISFAIPIDVANNV 286

Db 238 GAGPYDDIQTIDAANKSGGAPDLSGEVIGINTAIFSPSGSVGIAFAPISSTAKOV 297
QY 287 AEOLKNTKGVORGLGVIIQEVSYGLAOSFGLDKASGALLIAKLIPSPAPRAGIQAQDIY 346
Db 298 VDQLTKKGSVERGWIQVQIOPVTKDIAASIGLAEEKGAIYASPODDGPAKAKIKKGDVI 357
QY 347 LSLDGEIRSSGDLVVMVCAITPGKEVSLGWWR--KGEITIK-ALGNAAEITGASSTK 403
Db 358 TAVNGETVQDPEDLARKVANIAFGKAKALTWKRNKFAEEINVTIAMPMDDKSGSGQSDND 417
QY 404 DEAPYTEOQSGTFPSVESAGITLTQHTDSSGKHLVVYV--SDAARRAGLRHDEILAV 459
Db 418 NDGGGE-----TLDSTGLTVPSRD--GKGIVTVDDVPDSDAAR-GIRSDVIVSV 467

RESULT 15

ID 08YI32 PRELIMINARY; PRT; 524 AA.
AC 08YI32
DT 01-MAR-2002 (TRENBLrel. 20, Created)
01-MAR-2002 (TRENBLrel. 20, Last sequence update)
01-JUN-2002 (TRENBLrel. 21, Last annotation update)
Protease DO (EC 3.4.21.-).
BME10613.
GN Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
RX MEDLINE=20020109; PubMed=11756688;
RA Delvecchio V.G., Kapralov V., Redkar R.J., Patra G., Mujer C., Los T.,
Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,
Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
Selkov E., Elizer P.H., Hagius S., Overbeek R.,
Haseilhorn R., Kyrpides N., Overbeek R.,
"The genome sequence of the facultative intracellular pathogen
RT Brucella melitensis."
RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
DR EMBL; AE009503; AAL51794.1; -.
DR InterPro; IPR001478; PDZ.
DR InterPro; IPR001940; Protease2C.
DR InterPro; IPR001254; Ser_protease_Try.
DR Pfam; PF00595; PDZ; 2.
DR Pfam; PF00089; trypsin; 1.
DR PRINTS; PR00834; PROTEASES2C.
DR SMART; SM00228; PDZ; 2.
PROSITE; PS50106; PDZ; 2.
Hydrolase; Complete proteome.
SEQUENCE 524 AA; 55273 MW; 15B98E2452223658 CRC64;

Query Match 29.0%; Score 683; DB 16; Length 524;
Best Local Similarity 40.5%; Pred. No. 8.9e-34;
Matches 161; Conservative 67; Mismatches 152; Indels 18; Gaps 7;

QY 48 GSVSMILP-----DPAQLVQSEGPVNNIQAAPARTONGSGNAETSDPLADSDP 98
Db 26 GTVMNGTPTALAAQGPASVADLAEGLDVAVNISTQYVK-DDGEGDGPVMPQVPEGSP 84
QY 99 FYEFFKRLVPMNPEIPEBEADGGINFGSGFTI-SKNGYILTNTYVAVGWSIKVLLNDK 157
Db 85 FOEFFKDFP-NDKDAQGGDSRKVQSLGSGFTIIDAKGYITVNNHVIADADEIEVNFNDG 143
QY 158 REYTRAKLIGSDVQSVALLIKIDATE-ELFVYKIGNPKNLKPGEWVAIGAPPGFDSVTA 216
Db 144 SKLAEELVQKOTKTLAILKVPDSKRLAVHFGNSEKARIGDVIILGMPFGLGIVTA 203
QY 217 GIVSAKGRSLPMSSTYPTFIQTVAINPNSGGPLFVLKGVGVGINSQIYRSRSGGFMGISF 276
Db 204 GIIARAKDINSGRPYDDFIQTIDPAINRNGSGPLFMDGKVGIGINTAITSPSGSGIGIGF 263
QY 277 AIPDIVANNVAEQLKNTKGVORGLGVIIQEVSYGLAOSFGLDKASGALLIAKLIPSPAR 316

Db 264 AIPAEWMAAGVTDQKERGEVRRGWLGVRLQPTVDIAQSLGKTKGALLIAGLIENSGVD 323
QY 337 RAGLQAGDIYLSLDGEIRSSGDLVVMVCAITPGKEVSLGWWRKGEITIKAKLGNAAEH 396
Db 324 NKALHAGDVIRPDKRVDPTRADLPRLVAERPVGKEVEIVIRQGAETIKKVLGCLVE- 382
QY 397 TGASSKTDEAPYTEOQSGTFPSVESAGITLTQHTDSSGK 434
Db 383 ---DDKSTPEAVEDQVPAPDDGEQPGARQET-PDKSDK 416

Search completed: July 11, 2003, 10:33:11
Job time : 87 secs

